

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 10.9751 Seconds  
(without alignments)  
7635.278 Million cell updates/sec

Title: US-10-001-857-41  
Perfect score: 151  
Sequence: 1 cgcgccgggcaggtagcctaa.....gcaaatctctcaagaatatg 151

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	31	20.5	1689	4	US-09-489-039A-840
C 2	30.2	20.0	4421	4	US-08-952-674-1
C 3	29.4	19.5	2348	3	US-08-480-173A-42
C 4	29.4	19.5	2348	3	US-08-484-408A-42
C 5	29.4	19.5	3182	4	US-08-890-735C-1
C 6	29.4	19.5	3182	4	US-10-104-966-14
C 7	29.4	19.5	5618	3	US-08-799-569-1
C 8	29.4	19.5	5618	4	US-09-570-546-1
C 9	29.4	19.5	5618	4	US-09-146-072A-1
C 10	29	19.2	176373	3	US-09-128-155-17
C 11	28.6	18.9	3215	4	US-09-719-528A-1
C 12	28	18.5	786431	4	US-09-751-389-3
C 13	27.8	18.4	49312	4	US-09-671-317-485
C 14	27.4	18.1	1664976	4	US-08-916-421B-1
C 15	27.2	18.0	595	4	US-08-976-259-87
C 16	26.8	17.7	1740	4	US-09-328-352-999
C 17	26.8	17.7	168575	4	US-09-426-290-1
C 18	26.6	17.6	501	4	US-09-621-976-2040
C 19	26.6	17.6	2089	4	US-09-620-312D-303
C 20	26.6	17.6	3069	4	US-09-514-907A-5
C 21	26.6	17.6	3069	4	US-09-896-994-5
C 22	26.6	17.6	392000	4	US-10-027-983-11
C 23	26.6	17.6	580073	4	US-08-545-528D-1
C 24	26.4	17.5	577	3	US-09-328-111-606
C 25	26.4	17.5	1671	4	US-09-545-686-1
C 26	26.2	17.4	2111	4	US-10-096-571-11
C 27	26.2	17.4	2111	4	US-10-096-571-13

C 28	26.2	17.4	2841	2	US-08-452-075-1	Sequence 1, Appli
C 29	26.2	17.4	2841	3	US-09-231-061-1	Sequence 1, Appli
C 30	26.2	17.4	2841	4	US-09-011-762-5	Sequence 5, Appli
C 31	26	17.2	312	4	US-09-313-294A-7377	Sequence 7377, Ap
C 32	26	17.2	779	4	US-08-956-171E-1019	Sequence 1019, Ap
C 33	26	17.2	1449	4	US-08-956-171E-592	Sequence 592, App
C 34	26	17.2	41708	4	US-09-470-512A-3	Sequence 3, Appli
C 35	26	17.2	392000	4	US-10-027-983-11	Sequence 11, Appl
C 36	26	17.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 37	26	17.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
C 38	25.8	17.1	618	4	US-09-489-039A-1035	Sequence 1035, Ap
C 39	25.8	17.1	5108	4	US-09-306-446C-8	Sequence 8, Appli
C 40	25.8	17.1	5889	1	US-08-463-092B-5	Sequence 5, Appli
C 41	25.8	17.1	5889	2	US-08-462-109A-5	Sequence 5, Appli
C 42	25.8	17.1	5889	2	US-08-460-907B-5	Sequence 5, Appli
C 43	25.8	17.1	5889	3	US-08-463-179A-5	Sequence 5, Appli
C 44	25.8	17.1	5889	3	US-08-461-384B-5	Sequence 5, Appli
C 45	25.8	17.1	580073	4	US-08-545-528D-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-489-039A-840/c  
; Sequence 840, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 840  
; LENGTH: 1689  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-840

Query Match 20.5%; Score 31; DB 4; Length 1689;  
Best Local Similarity 68.3%; Pred. No. 0.24;  
Matches 43; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY	28	AAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTC	87
Db	1607	ACATGAAGCAGGTATGGCCCTGGCTCGAAAGCATGGTGTGCGAGTCCGCTTCAACCTGG	1548
QY	88	TCT 90	
Db	1547	TCT 1545	

RESULT 2  
US-08-952-674-1/c  
; Sequence 1, Application US/08952674  
; Patent No. 6623951  
; GENERAL INFORMATION:  
; APPLICANT: Hofschneider, Peter  
; APPLICANT: Habenberger, Peter  
; APPLICANT: Ludwig, Weiss  
; TITLE OF INVENTION: HBV Vectors and Cells for Producing Same  
; FILE REFERENCE: HBV Vectors  
; CURRENT APPLICATION NUMBER: US/08/952,674  
; CURRENT FILING DATE: 1998-03-02  
; EARLIER APPLICATION NUMBER: PCT/DE96/00807  
; EARLIER FILING DATE: 1996-05-09  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1

; LENGTH: 4421  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: pHBV/V1  
US-08-952-674-1

Query Match 20.0%; Score 30.2; DB 4; Length 4421;  
Best Local Similarity 53.9%; Pred. No. 0.71;  
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 18 TAAACAGGCCAAATGTTGCCTTTTGGGTTTCCTGTTTCAACAGCATGGTGTGAAGCGCCG 77  
Db 4177 TGAACAGACCAATTATGCCTACAGCCTCCTAGTACAAAGACCTTTAACTTAATCTCTC 4118  
QY 78 ATCAACCTTCTGCTTATTAAATAAAATGTCATAAACTCATCTGCAAGGTGG 132  
Db 4117 CCCCACTCTCCAGTCTTAAACAACAGTCTTTGAAGTATGCCTCAAGGTCG 4063

RESULT 3  
US-08-480-173A-42/c  
; Sequence 42, Application US/08480173A  
; Patent No. 6072049  
; GENERAL INFORMATION:  
; APPLICANT: Thoma, Hans A  
; TITLE OF INVENTION: HEPATITIS B SURFACE ANTIGEN VACCINE  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Popovich & Wiles, P.A.  
; STREET: 80 S. 8th Street, Suite 1902  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,173A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Popovich, Thomas E  
; REGISTRATION NUMBER: 30,099  
; REFERENCE/DOCKET NUMBER: MED1003USD4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-334-8991  
; TELEFAX: 612-334-8994  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2348 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 12..14  
; OTHER INFORMATION: /note= "S1 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 336..338  
; OTHER INFORMATION: /note= "S2 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 508..510  
; OTHER INFORMATION: /note= "S start codon"

US-08-484-408A-42/c  
; Sequence 42, Application US/08484408A  
; Patent No. 6117653  
; GENERAL INFORMATION:  
; APPLICANT: Thoma, Hans A  
; TITLE OF INVENTION: HEPATITIS B SURFACE ANTIGEN VACCINE  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Popovich & Wiles, P.A.  
; STREET: 80 S. 8th Street, Suite 1902  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,408A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Popovich, Thomas E  
; REGISTRATION NUMBER: 30,099  
; REFERENCE/DOCKET NUMBER: MED1003USD4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-334-8991  
; TELEFAX: 612-334-8994  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2348 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 12..14  
; OTHER INFORMATION: /note= "S1 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 336..338  
; OTHER INFORMATION: /note= "S2 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 508..510  
; OTHER INFORMATION: /note= "S start codon"

US-08-484-408A-42/c  
; Sequence 42, Application US/08484408A  
; Patent No. 6117653  
; GENERAL INFORMATION:  
; APPLICANT: Thoma, Hans A  
; TITLE OF INVENTION: HEPATITIS B SURFACE ANTIGEN VACCINE  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Popovich & Wiles, P.A.  
; STREET: 80 S. 8th Street, Suite 1902  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,408A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Popovich, Thomas E  
; REGISTRATION NUMBER: 30,099  
; REFERENCE/DOCKET NUMBER: MED1003USD4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-334-8991  
; TELEFAX: 612-334-8994  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2348 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 12..14  
; OTHER INFORMATION: /note= "S1 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 336..338  
; OTHER INFORMATION: /note= "S2 start codon"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 508..510  
; OTHER INFORMATION: /note= "S start codon"

Query Match 19.5%; Score 29.4; DB 3; Length 2348;  
Best Local Similarity 54.1%; Pred. No. 1;  
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTTGCCTTTTGGGTTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCA 81  
Db 2154 CAGACCAATTTATGCCTACAGCCTCCTAGTACAAAGACCTTTAACTTAATCTCTCCCC 2095

QY 82 ACCTTCTCTGCCTATTAAATAAAATGTCATAAACTCATCTGCAAGGTGG 132  
Db 2094 AACTCTCTCCAGTCTTTAAACAACAGTCTTTGAAGTATGCCTCAAGGTGG 2044

Query Match 19.5%; Score 29.4; DB 3; Length 2348;  
Best Local Similarity 54.1%; Pred. No. 1;  
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTTGCCTTTTGGGTTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCA 81  
Db 2154 CAGACCAATTTATGCCTACAGCCTCCTAGTACAAAGACCTTTAACTTAATCTCTCCCC 2095

QY 82 ACCTTCTCTGCCTATTAAATAAAATGTCATAAACTCATCTGCAAGGTGG 132  
Db 2094 AACTCTCTCCAGTCTTTAAACAACAGTCTTTGAAGTATGCCTCAAGGTGG 2044

Query Match 19.5%; Score 29.4; DB 3; Length 2348;  
Best Local Similarity 54.1%; Pred. No. 1;  
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 CAGGCCAAATGTTGCCTTTTGGGTTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCA 81  
Db 2154 CAGACCAATTTATGCCTACAGCCTCCTAGTACAAAGACCTTTAACTTAATCTCTCCCC 2095

QY 82 ACCTTCTCTGCCTATTAAATAAAATGTCATAAACTCATCTGCAAGGTGG 132  
Db 2094 AACTCTCTCCAGTCTTTAAACAACAGTCTTTGAAGTATGCCTCAAGGTGG 2044







```
; Patent No. 6558675
; GENERAL INFORMATION:
; APPLICANT: Oon, Chong Jin
; Lim, Gek Keow
; Zhao, Yi
; Chen, Wei Ning
; TITLE OF INVENTION: A MUTANT HUMAN HEPATITIS B VIRAL STRAIN AND
; USES THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/719,528A
; FILING DATE: 30-Apr-2001
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SG98/00046
; FILING DATE: 19-JAN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mass, Clifford J.
; REGISTRATION NUMBER: 30,086
; REFERENCE/DOCKET NUMBER: U-013109-7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 708-1800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-719-528A-1

Query Match      18.9%; Score 28.6; DB 4; Length 3215;
Best Local Similarity 53.0%; Pred. No. 2.3;
Matches 61; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY      18  TAAACAGGCCAAATGTTGCCTTTGGGGTTCTGTTCACACAGCATGGTGTGAAGCGCCG 77
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1805  TGAACAGACCAATTATGCCTACAGCCTCCTAGTACATAAATCTTTAACTAATCTCCTC 1746

QY      78  ATCAACCTTCTCTGCCTATTAAAAATAAAATGTCATAAACTCATCTGCAAGGTGG 132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1745  CCCCAACTCCTCCAGTCTTTAAACACACACAGTCTTTGAAGTATGCCTCAAGGTCG 1691

RESULT 12
US-09-751-389-3
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match      18.5%; Score 28; DB 4; Length 786431;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      44  GTTCCTGTTTCAACAGCATGGTGTGAAGCGCCGCATCAACCTTCTCTGCCTATTAAAAATA 103
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      157775  GTTACTGTTACAGGAGCATATTCCTAAGTTAGTTTTCATCTTGCTGCTTAAAGCTA 157834

RESULT 13
US-09-671-317-485
; Sequence 485, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 485
; LENGTH: 49312
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5466..7466
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 7467..7725
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 20256..20355
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 36905..36975
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 45167..45248
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 45728..45965
; OTHER INFORMATION: exon 5
; NAME/KEY: misc_feature
; LOCATION: 45966..49312
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 7564
; OTHER INFORMATION: 10-286-289 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 7619
; OTHER INFORMATION: 10-286-345 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 7649
```

OTHER INFORMATION: 10-286-375 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 17258  
OTHER INFORMATION: 12-425-57 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 21590  
OTHER INFORMATION: 12-421-135 : insertion of T  
NAME/KEY: allele  
LOCATION: 21595  
OTHER INFORMATION: 12-421-140 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 36971  
OTHER INFORMATION: 10-523-232 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 45214  
OTHER INFORMATION: 10-289-201 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 45741  
OTHER INFORMATION: 10-290-37 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 46029  
OTHER INFORMATION: 10-290-326 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 46032  
OTHER INFORMATION: 10-290-328 : deletion of G  
NAME/KEY: primer\_bind  
LOCATION: 7276..7294  
OTHER INFORMATION: 10-286.pu  
NAME/KEY: primer\_bind  
LOCATION: 7676..7694  
OTHER INFORMATION: 10-286.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 16839..16856  
OTHER INFORMATION: 12-425.rp  
NAME/KEY: primer\_bind  
LOCATION: 17297..17314  
OTHER INFORMATION: 12-425.pu complement  
NAME/KEY: primer\_bind  
LOCATION: 21456..21474  
OTHER INFORMATION: 12-421.pu  
NAME/KEY: primer\_bind  
LOCATION: 21886..21906  
OTHER INFORMATION: 12-421.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 36740..36758  
OTHER INFORMATION: 10-523.pu  
NAME/KEY: primer\_bind  
LOCATION: 36997..37015  
OTHER INFORMATION: 10-523.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 45020..45037  
OTHER INFORMATION: 10-289.pu  
NAME/KEY: primer\_bind  
LOCATION: 45413..45432  
OTHER INFORMATION: 10-289.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 45705..45724  
OTHER INFORMATION: 10-290.pu  
NAME/KEY: primer\_bind  
LOCATION: 46104..46123  
OTHER INFORMATION: 10-290.rp complement  
NAME/KEY: primer\_bind  
LOCATION: 7545..7563  
OTHER INFORMATION: 10-286-289.mis  
NAME/KEY: primer\_bind  
LOCATION: 7565..7583  
OTHER INFORMATION: 10-286-289.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 7600..7618  
OTHER INFORMATION: 10-286-345.mis  
NAME/KEY: primer\_bind  
LOCATION: 7620..7638  
OTHER INFORMATION: 10-286-345.mis complement

NAME/KEY: primer\_bind  
LOCATION: 7630..7648  
OTHER INFORMATION: 10-286-375.mis  
NAME/KEY: primer\_bind  
LOCATION: 7650..7668  
OTHER INFORMATION: 10-286-375.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 17239..17257  
OTHER INFORMATION: 12-425-57.mis  
NAME/KEY: primer\_bind  
LOCATION: 17259..17277  
OTHER INFORMATION: 12-425-57.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 21576..21594  
OTHER INFORMATION: 12-421-140.mis  
NAME/KEY: primer\_bind  
LOCATION: 21596..21614  
OTHER INFORMATION: 12-421-140.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 36952..36970  
OTHER INFORMATION: 10-523-232.mis  
NAME/KEY: primer\_bind  
LOCATION: 36972..36990  
OTHER INFORMATION: 10-523-232.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 45195..45213  
OTHER INFORMATION: 10-289-201.mis  
NAME/KEY: primer\_bind  
LOCATION: 45215..45233  
OTHER INFORMATION: 10-289-201.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 45722..45740  
OTHER INFORMATION: 10-290-37.mis  
NAME/KEY: primer\_bind  
LOCATION: 45742..45760  
OTHER INFORMATION: 10-290-37.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 46010..46028  
OTHER INFORMATION: 10-290-326.mis  
NAME/KEY: primer\_bind  
LOCATION: 46030..46048  
OTHER INFORMATION: 10-290-326.mis complement  
NAME/KEY: misc\_binding  
LOCATION: 7552..7576  
OTHER INFORMATION: 10-286-289.probe  
NAME/KEY: misc\_binding  
LOCATION: 7607..7631  
OTHER INFORMATION: 10-286-345.probe  
NAME/KEY: misc\_binding  
LOCATION: 7637..7661  
OTHER INFORMATION: 10-286-375.probe  
NAME/KEY: misc\_binding  
LOCATION: 17246..17270  
OTHER INFORMATION: 12-425-57.probe  
NAME/KEY: misc\_binding  
LOCATION: 21583..21607  
OTHER INFORMATION: 12-421-140.probe  
NAME/KEY: misc\_binding  
LOCATION: 36959..36983  
OTHER INFORMATION: 10-523-232.probe  
NAME/KEY: misc\_binding  
LOCATION: 45202..45226  
OTHER INFORMATION: 10-289-201.probe  
NAME/KEY: misc\_binding  
LOCATION: 45729..45753  
OTHER INFORMATION: 10-290-37.probe  
NAME/KEY: misc\_binding  
LOCATION: 46017..46041  
OTHER INFORMATION: 10-290-326.probe  
US-09-671-317-485

Query Match 18.4%; Score 27.8; DB 4; Length 49312;  
Best Local Similarity 65.1%; Pred. No. 16;



```
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 52 TTCAACAGCATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCA 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23699 TTCAAGAGCAGCCTGGGCAACATAGCAAGAGCCCTCTCTCTATTAAAAAAATTAAA 23758
QY 112 TAA 114
Db 23759 AAA 23761

RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
```

NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854)..(1664855)  
OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1

Query Match 18.1%; Score 27.4; DB 4; Length 1664976;  
Best Local Similarity 55.9%; Pred. No. 91;  
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
Qy 28 AAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTC 87  
Db 640752 AAATGAAGAAATTTGGGTAACTCCTGCTATACCATTTGTTGTCGATATTTCTATCTTA 640693  
Qy 88 TCTGCCTATTAAATAAAATGTCATAAACTCAT 120  
Db 640692 TTGCTAACATCAATTATAGGTGATATAAATCAT 640650

RESULT 15  
US-08-976-259-87  
Sequence 87, Application US/08976259  
Patent No. 6316609  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Choi, Gil H.  
APPLICANT: Welch, Rodney A.  
TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli  
Patent No. 6316609  
NUMBER OF SEQUENCES: 142  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave, N.W., Suite 600  
CITY: Washington  
STATE: DC

COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/976,259  
FILING DATE: Herewith  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,626 AND US 60/061,953  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 87:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-976-259-87

Query Match 18.0%; Score 27.2; DB 4; Length 595;  
Best Local Similarity 67.9%; Pred. No. 3.5;  
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
Qy 95 ATTAAATAAAATGTCAATAAACTCATCCTGCAAGGTGGCAATTCCTCAAGAATAT 150  
Db 368 ATTAAAGCAACATGCAGATTCTTTCATCCAAACCCGCAACTTCTCTCATAAATAT 423  
Search completed: August 10, 2004, 21:58:15  
Job time : 14.9751 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 601.675 Seconds  
(without alignments)  
10877.634 Million cell updates/sec

Title: US-10-001-857-41  
Perfect score: 151  
Sequence: 1 ccgccgggcaggtagcctaa.....gcaaatctcctaagaatatg 151

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	151	100.0	151	6	AX523453	AX523453 Sequence
C 2	140.4	93.0	3096	6	AX523454	AX523454 Sequence
C 3	106.4	70.5	160115	9	AL161447	AL161447 Human DNA
C 4	106.4	70.5	320902	2	AL161456	AL161456 Homo sapi
C 5	84.8	56.2	2488	6	AX537452	AX537452 Sequence
C 6	83.8	55.5	815	6	AX523523	AX523523 Sequence
C 7	83.2	55.1	1839	9	AK025266	AK025266 Homo sapi
C 8	72.6	48.1	281804	2	AC134869	AC134869 Mus muscu
C 9	70.6	46.8	253611	2	AC111473	AC111473 Rattus no
C 10	70.6	46.8	271861	2	AC108541	AC108541 Rattus no
C 11	67.4	44.6	2552	10	AY102701	AY102701 Mus muscu
C 12	67.4	44.6	2594	10	BC056435	BC056435 Mus muscu
C 13	63.2	41.9	2543	10	AF272892	AF272892 Rattus no
C 14	53	35.1	2467	5	BC053286	BC053286 Danio rer
C 15	53	35.1	147876	2	AC102010	AC102010 Mus muscu
C 16	36.6	24.2	3215	14	AB074755	AB074755 Hepatitis
C 17	35.8	23.7	70276	9	AC073095	AC073095 Homo sapi
C 18	35	23.2	266	14	AF276569	AF276569 Hepatitis
C 19	35	23.2	266	14	AF276570	AF276570 Hepatitis
C 20	35	23.2	266	14	AF276571	AF276571 Hepatitis
C 21	35	23.2	266	14	AF276572	AF276572 Hepatitis
C 22	35	23.2	266	14	AF276573	AF276573 Hepatitis
C 23	35	23.2	266	14	AF276574	AF276574 Hepatitis
C 24	35	23.2	266	14	AF276575	AF276575 Hepatitis
C 25	35	23.2	266	14	AF276576	AF276576 Hepatitis
C 26	35	23.2	465	14	AF378791	AF378791 Hepatitis
C 27	35	23.2	3202	14	AY217374	AY217374 Hepatitis
C 28	35	23.2	3206	14	AB112408	AB112408 Hepatitis
C 29	35	23.2	3215	14	AB112471	AB112471 Hepatitis
C 30	35	23.2	3215	14	AF068756	AF068756 Hepatitis
C 31	35	23.2	3215	14	AF223960	AF223960 Hepatitis
C 32	34.8	23.0	172542	2	AC146320	AC146320 Papio anu
C 33	34.8	23.0	181792	9	AC093126	AC093126 Papio anu
C 34	34.6	22.9	506	14	AF325702	AF325702 Hepatitis
C 35	34.6	22.9	512	14	AF325703	AF325703 Hepatitis
C 36	34.6	22.9	25400	3	CEF59A2	Z34801 Caenorhabdi
C 37	34.2	22.6	159994	2	AC073471	AC073471 Homo sapi
C 38	34	22.5	466	14	AY251136	AY251136 Hepatitis
C 39	33.8	22.4	5751	6	BD181579	BD181579 Novel gen
C 40	33.8	22.4	5751	9	AB058692	AB058692 Homo sapi
C 41	33.8	22.4	23631	9	HSV311G7	Z69304 Human DNA s
C 42	33.8	22.4	161078	9	AP003500	AP003500 Homo sapi
C 43	33.8	22.4	161286	2	AC025120	AC025120 Homo sapi
C 44	33.8	22.4	168467	10	AC123798	AC123798 Mus muscu
C 45	33.8	22.4	180303	9	AP003782	AP003782 Homo sapi

ALIGNMENTS

RESULT 1  
AX523453  
LOCUS AX523453 151 bp DNA linear PAT 24-OCT-2002  
DEFINITION Sequence 41 from Patent WO02064788.

ACCESSION AX523453

VERSION AX523453.1 GI:24412349

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.  
TITLE Compositions and methods relating to lung specific genes and proteins

JOURNAL Patent: WO 02064788-A 41 22-AUG-2002;  
Diadexus, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..151  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 151; DB 6; Length 151;  
Best Local Similarity 100.0%; Pred. No. 7.6e-38;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 60  
|||||  
Db 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGC 60  
|||||

QY 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATAAATGTCATAAACTCAT 120  
|||||  
Db 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATAAATGTCATAAACTCAT 120  
|||||

QY 121 CCTGCAAGTGGCAATTCCTCAAGAAATG 151  
|||||  
Db 121 CCTGCAAGTGGCAATTCCTCAAGAAATG 151  
|||||

RESULT 2  
AX523454/c  
LOCUS AX523454 3096 bp DNA linear PAT 24-OCT-2002  
DEFINITION Sequence 42 from Patent WO02064788.  
ACCESSION AX523454  
VERSION AX523454.1 GI:24412350  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.  
Compositions and methods relating to lung specific genes and  
proteins  
Patent: WO 02064788-A 42 22-AUG-2002;  
Diadexus, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..3096  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 93.0%; Score 140.4; DB 6; Length 3096;  
Best Local Similarity 99.3%; Pred. No. 2.9e-34;  
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69  
|||||  
Db 2052 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 1993  
|||||

QY 70 AGCGCGCATCAACCTTCTGCTTATTAAATAAATGTCATAAACTCATCTGCAAGG 129  
|||||  
Db 1992 AGCGCTGCATCAACCTTCTGCTTATTAAATAAATGTCATAAACTCATCTGCAAGG 1933  
|||||

QY 130 TGGCAAAATTCCTCAAGAAATG 151  
|||||  
Db 1932 TGGCAAAATTCCTCAAGAAATG 1911  
|||||

RESULT 3  
AL161447/c  
LOCUS AL161447 160115 bp DNA linear PRI 16-MAR-2002  
DEFINITION Human DNA sequence from clone RP11-379P1 on chromosome 9, complete  
sequence.  
ACCESSION AL161447

AL161447.20 GI:19572746  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 160115)  
Johnson,C.  
Direct Submission  
Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Mar 21, 2002 this sequence version replaced gi:15787718.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-379P1 is from the library RPCI-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
This sequence is the entire insert of clone RP11-379P1 The true  
left end of clone RP11-280P22 is at 137145 in this sequence. The  
true right end of clone RP11-65C15 is at 38497 in this sequence.  
Location/Qualifiers  
1..160115  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-379P1"  
/clone\_lib="RPCI-11.2"  
complement(32505..32575)  
/note="Sequence from uni-directional dGTP big dye  
terminator reads only."  
74372..74537  
/note="Single clone region. Assembly confirmed by  
restriction digest data."  
74706..74920  
/note="Single clone region. Assembly confirmed by  
restriction digest data."  
complement(74706..74920)  
/note="Sequence from uni-directional dGTP big dye  
terminator reads only."  
74921  
/note="Tandem repeat. Forced join. Gap size estimated to  
be approximately 200bp by restriction digest data."  
129108  
/note="Tandem repeat. Forced join. Gap size estimated to  
be approximately 300bp by restriction digest data."

ORIGIN  
Query Match 70.5%; Score 106.4; DB 9; Length 160115;  
Best Local Similarity 99.1%; Pred. No. 4.6e-23;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





```
* 163498 166241: contig of 2744 bp in length
* 166242 166341: gap of 100 bp
* 166342 168447: contig of 2106 bp in length
* 168447 168547: gap of 100 bp
* 168548 171536: contig of 2989 bp in length
* 171537 171636: gap of 100 bp
* 171637 175769: contig of 4133 bp in length
* 175770 175869: gap of 100 bp
* 175870 183977: contig of 8108 bp in length
* 183978 184077: gap of 100 bp
* 184078 186608: contig of 2531 bp in length
* 186609 186708: gap of 100 bp
* 186709 190121: contig of 3413 bp in length
* 190122 190221: gap of 100 bp
* 190222 197833: contig of 7612 bp in length
* 197834 197933: gap of 100 bp
* 197934 200266: contig of 2333 bp in length
* 200267 200366: gap of 100 bp
* 200367 203220: contig of 2854 bp in length
* 203221 203320: gap of 100 bp
* 203321 208956: contig of 5636 bp in length
* 208957 209056: gap of 100 bp
* 209057 211798: contig of 2742 bp in length
* 211799 211898: gap of 100 bp
* 211899 214088: contig of 2190 bp in length
* 214089 214188: gap of 100 bp
* 214189 217324: contig of 3136 bp in length
* 217325 217424: gap of 100 bp
* 217425 220661: contig of 3237 bp in length
* 220662 220761: gap of 100 bp
* 220762 225678: contig of 4917 bp in length
* 225679 225778: gap of 100 bp
* 225779 228798: contig of 3020 bp in length
* 228799 228898: gap of 100 bp
* 228899 231289: contig of 2391 bp in length
* 231290 231389: gap of 100 bp
* 231390 235056: contig of 3667 bp in length
* 235057 235156: gap of 100 bp
* 235157 238252: contig of 3096 bp in length
* 238253 238352: gap of 100 bp
* 238353 241060: contig of 2708 bp in length
* 241061 241160: gap of 100 bp
* 241161 244514: contig of 3354 bp in length
* 244515 244614: gap of 100 bp
* 244615 247447: contig of 2833 bp in length
* 247448 247547: gap of 100 bp
* 247548 252282: contig of 4735 bp in length
* 252283 252382: gap of 100 bp
* 252383 254818: contig of 2436 bp in length
* 254819 254918: gap of 100 bp
* 254919 257094: contig of 2176 bp in length
* 257095 257194: gap of 100 bp
* 257195 259885: contig of 2691 bp in length
* 259886 259985: gap of 100 bp
* 259986 262579: contig of 2594 bp in length
* 262580 262679: gap of 100 bp
* 262680 267289: contig of 4610 bp in length
* 267290 267389: gap of 100 bp
* 267390 271789: contig of 4400 bp in length
* 271790 271889: gap of 100 bp
* 271890 280376: contig of 8487 bp in length
* 280377 280476: gap of 100 bp
* 280477 286094: contig of 5618 bp in length
* 286095 286194: gap of 100 bp
* 286195 293197: contig of 7003 bp in length
* 293198 293297: gap of 100 bp

Query Match      70.5%; Score 106.4; DB 2; Length 320902;
Best Local Similarity 99.1%; Pred. No. 5e-23;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
|||||
```

```
Db 261433 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 261374

QY 70 AGCGCCGCATCAACCTTCTCTGCTTATTAAATAAAATGTCATAAACT 117
|||||

Db 261373 AGCGCTGCATCAACCTTCTCTGCTTATTAAATAAAATGTCATAAACT 261326

RESULT 5
AX537452/c
LOCUS      AX537452      2488 bp      DNA      linear      PAT 23-NOV-2002
DEFINITION Sequence 23 from Patent WO02070709.
ACCESSION  AX537452
VERSION     AX537452.1  GI:25269238
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Lal,P.G., Baughn,M.R., Yao,M.G., Walia,N.K., Elliot,V.S., Xu,Y.,
Honchell,C.D., Yue,H., Ding,L., Gietzen,K.J., Ison,C.H., Lu,D.A.,
Hafalia,A.J., Ghandi,A.R., Thangavelu,K., Sanjanwala,M.M.,
Tang,Y.T., Ramkumar,J., Griffin,J.A., Swarnaker,A., Azimzai,Y.,
Sapperstein,S.K., Burford,N., Lee,E.A., Lu,Y., Tran,U.K. and
Marquis,J.P.
TITLE       Molecules for disease detection and treatment
JOURNAL     Patent: WO 02070709-A 23 12-SEP-2002;
Incyte Genomics, Inc. (US)
FEATURES    Location/Qualifiers
            source
            1..2488
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                /note="Incyte ID No: 7111920CB1"

ORIGIN
Query Match      56.2%; Score 84.8; DB 6; Length 2488;
Best Local Similarity 88.5%; Pred. No. 2.6e-16;
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
|||||

Db 1554 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 1495
|||||

QY 70 AGCGCCGCATCAACCTTCTCTGCTTATTAAATAAAATGTCATA 113
|||||

Db 1494 AGCGCTGCATCAACCTTCTCTGCTTCATCCTGCAAGGTGGCAAA 1451
|||||

RESULT 6
AX523523
LOCUS      AX523523      815 bp      DNA      linear      PAT 24-OCT-2002
DEFINITION Sequence 111 from Patent WO02064788.
ACCESSION  AX523523
VERSION     AX523523.1  GI:24412419
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
TITLE       Compositions and methods relating to lung specific genes and
proteins
JOURNAL     Patent: WO 02064788-A 111 22-AUG-2002;
Diadexus, Inc. (US)
FEATURES    Location/Qualifiers
            source
            1..815
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"

ORIGIN
```

Query Match 55.5%; Score 83.8; DB 6; Length 815;  
Best Local Similarity 88.3%; Pred. No. 4.8e-16;  
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 70  
|||||  
Db 87 AGGTACTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 146  
|||||

QY 71 GCGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
|||||  
Db 147 GCGTGCATCAACCTTCTCTGCCTCATCCTGCAAGGTGGCAA 189  
|||||

RESULT 7  
AK025266/c  
LOCUS AK025266 1839 bp mRNA linear PRI 13-SEP-2003  
DEFINITION Homo sapiens cDNA: FLJ21613 fis, clone COL07381.  
ACCESSION AK025266  
VERSION AK025266.1 GI:10437744  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 1839)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
Direct Submission  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan. (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source  
1. .1839  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COL07381"  
/tissue type="colon"  
/clone\_lib="COL"  
/note="cloning vector pME18SFL3"  
134. .1714  
/note="unnamed protein product"  
/codon start=1  
/protein id="BAB15097.1"  
/db\_xref="GI:10437745"  
/translation="MVVKASVDDDDSGWELSMPEKMEKSNWNVDITQDFEEACRELK  
LGELHDKLFGLFEMSAIEMDPKMDGMIGNQVNRKVLNFEQAIKDGTIKIKDLTL  
PELIGMDTCFCCCLITWLEGHSLAQVFTCLYIHNPDFIEDPAKAFALGILKICDIA  
REKVNKAAVFEEEDFQSMTYGFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGE  
RDPVELEHQRCIAVFSRVKFTRVLLTVLIAFTKETSAAVEAQKLMVQAADLLSAIH  
NSLHHGIQAQNDTTKGDHPIMMGFEPLVNQRLLPPTFPFYAKIKREEMVNYFARLID  
RIKTVCVYVNLTLNLHCILDFFCFSEQSPCVLSRLQTTFLVDNKKVFGTHLMQDMV  
KDALREFSVPVLSPKCYLYNNHQAQKDCIDSFVTHCVRPFCSLIQIHGHNRRQRDKL  
GHLEEFATLQDEAEKVDAALHTMLLKQEPQQRQLAWLGTWVLYHNLRLIMIQYLLSGF  
EELYSMHAILHLIVSL"

CDS  
ORIGIN  
Query Match 55.1%; Score 83.2; DB 9; Length 1839;

Best Local Similarity 87.5%; Pred. No. 8.2e-16;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69  
|||||  
Db 1605 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 1546  
|||||

QY 70 AGCGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
|||||  
Db 1545 AGCGTGCATCAACCTTCTCTGCCTCATCCTGCAAGGTGGCAA 1502  
|||||

RESULT 8  
AC134869/c  
LOCUS AC134869 281804 bp DNA linear HTG 01-MAR-2003  
DEFINITION Mus musculus chromosome UNK clone RP23-137C6, WORKING DRAFT  
SEQUENCE, 15 unordered pieces.  
ACCESSION AC134869  
VERSION AC134869.3 GI:28626922  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 281804)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE The sequence of Mus musculus clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 281804)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-OCT-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 281804)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
COMMENT On Mar 1, 2003 this sequence version replaced gi:27877434.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site:http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: M\_BA0137C06  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 278252 bases at least Q40  
Consensus quality: 279651 bases at least Q30  
Consensus quality: 280523 bases at least Q20  
Insert size: 171000; agarose-fp  
Insert size: 282542; sum-of-contigs  
Quality coverage: 15.67 in Q20 bases; agarose-fp  
Quality coverage: 10.45 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1074: contig of 1074 bp in length  
\* 1075 1174: gap of unknown length  
\* 1175 2325: contig of 1151 bp in length



\* 2326 2425: gap of unknown length  
\* 2426 4173: contig of 1748 bp in length  
\* 4174 4273: gap of unknown length  
\* 4274 6930: contig of 2657 bp in length  
\* 6931 7030: gap of unknown length  
\* 7031 10444: contig of 3414 bp in length  
\* 10445 10544: gap of unknown length  
\* 10545 14581: contig of 4037 bp in length  
\* 14582 14681: gap of unknown length  
\* 14682 17225: contig of 2544 bp in length  
\* 17226 17325: gap of unknown length  
\* 17326 22780: contig of 5455 bp in length  
\* 22781 22880: gap of unknown length  
\* 22881 34746: contig of 11866 bp in length  
\* 34747 34846: gap of unknown length  
\* 34847 51109: contig of 16263 bp in length  
\* 51110 51209: gap of unknown length  
\* 51210 66908: contig of 15699 bp in length  
\* 66909 67008: gap of unknown length  
\* 67009 85189: contig of 18181 bp in length  
\* 85190 85289: gap of unknown length  
\* 85290 119762: contig of 34473 bp in length  
\* 119763 119862: gap of unknown length  
\* 119863 190705: contig of 70843 bp in length  
\* 190706 190805: gap of unknown length  
\* 190806 281804: contig of 90999 bp in length.  
Location/Qualifiers  
1. .281804  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP23-137C6"  
1. .1074  
/note="assembly\_name:Contig10"  
1175. .2325  
/note="assembly\_name:Contig12"  
2426. .4173  
/note="assembly\_name:Contig27"  
4274. .6930  
/note="assembly\_name:Contig30"  
7031. .10444  
/note="assembly\_name:Contig31"  
10545. .14581  
/note="assembly\_name:Contig32"  
14682. .17225  
/note="assembly\_name:Contig33"  
17326. .22780  
/note="assembly\_name:Contig34"  
22881. .34746  
/note="assembly\_name:Contig35"  
34847. .51109  
/note="assembly\_name:Contig36"  
51210. .66908  
/note="assembly\_name:Contig37"  
67009. .85189  
/note="assembly\_name:Contig38"  
85290. .119762  
/note="assembly\_name:Contig39"  
119863. .190705  
/note="assembly\_name:Contig40"  
190806. .281804  
/note="assembly\_name:Contig41"

ORIGIN

Query Match 48.1%; Score 72.6; DB 2; Length 281804;  
Best Local Similarity 85.3%; Pred. No. 4.1e-12;  
Matches 81; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTCTCTGTTTCAACAGCATGGTGTGA 69  
Db 241454 CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCTGTGTTTCAACAACATAGTATGA 241395

QY 70 AGCGCCGCATCAACCTTCTCTGCCTATTAAATAA 104  
Db 241394 AGCGCTGCATCAACCTTCTCTGCCTGTCAAAACAA 241360

RESULT 9  
AC111473/c

LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

1 (bases 1 to 253611)  
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Soralle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 253611)  
Worley,K.C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE







/product="embryonic growth-associated protein EGAP"
 /protein\_id="AA052342.1"
 /db\_xref="GI:21539896"
 /translation="MVMKATVDDDASGWELGVPEKMEKSTSWVDITQDFEDACRELK
 LGELLHDKLFGLFEMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLSL
 PELIGIMDTCFCCCLITWLEGHSLAQVFTCLYIHNPDFIEDPAMKAFALGILKICDIA
 REKVNKAAFEEDFQSMTYGFKMANSVTDLRTGMLKDVEDDLQRRVKSTRSQGEE
 RDPEVELEHQQCCLAAPSRVKFTRVLLTLIIAFTKETSAAVEAQKLMVQAADLLSAIH
 TSLHHGIIQAQNGTTKGDPHPIMMGFEPVLNQRLPPTFPFYAKIIRKEEMVNYFSRLID
 RIKTVCENVNLPNLHCILDDFFCEFESEQSPCVLSRLQTTFLVDNKKVFGTHLMQDMV
 KDALRSFVSPVLPSPKCLYNNHQAQKDCIDSFVTHCVRFCSLVQIHGHNRAQRQDKL
 GHILEEFATLQDEAEKVDAAALHTMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGF
 ELELYSMHEYIYIYVYLSEFLYAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKKK
 KKVRLPSREITMSQAYQNMCAGMFKTMVAFMDGKVRKPKFELDSEQVRYEHRFAPFN
 SVMTPPPVHYLQFKEMSDLNKYSPPPQSPELYVAASKHFQQTKMILENIPNPDREVSR
 ILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSVHKYFPVVKLV"

ORIGIN

Query Match 44.6%; Score 67.4; DB 10; Length 2552;
 Best Local Similarity 87.1%; Pred. No. 1.1e-10;
 Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1526 CAGGTTCTAAGCAGGCCAGATGTGTCTCTGAGGCTCCTGTTTCAACAACATAGTATGA 1467

QY 70 AGCGCGCATCAACCTTCTCTGCCT 94
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1466 AGCGTGCATCAACCTTCTCTGCCT 1442

RESULT 12
 BC056435/c
 LOCUS
 DEFINITION Mus musculus RIKEN cdna C030004C14 gene, mRNA (cdna clone MGC:66717
 IMAGE:5707801), complete cds.

ACCESSION BC056435
 VERSION BC056435.1 GI:33989782
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2594)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE
 human and mouse cdna sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE
 PUBMED
 12477932

2 (bases 1 to 2594)
 Strausberg,R.
 Direct Submission
 Submitted (14-AUG-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
 COMMENT
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAK Plate: 124 Row: k Column: 11
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 21644570.

FEATURES
 Location/Qualifiers
 1..2594
 /organism="Mus musculus"
 /mol\_type="mRNA"
 /strain="C57BL/6"
 /db\_xref="taxon:10090"
 /clone="MGC:66717 IMAGE:5707801"
 /tissue\_type="Brain, mouse 15.5 dpc"
 /clone\_lib="NIH BMAP\_EV0"
 /lab\_host="DH10B"
 /note="Vector: pYX-ASC"

1..2594
 /gene="C030004C14Rik"
 /db\_xref="LocusID:78689"
 /db\_xref="MGI:1925939"
 42..2219
 /codon\_start=1
 /product="embryonic growth-associated"
 /protein\_id="AAH56435.1"
 /db\_xref="GI:33989783"
 /db\_xref="LocusID:78689"
 /translation="MVMKATVDDDASGWELGVPEKMEKSTSWVDITQDFEDACRELK
 LGELLHDKLFGLFEMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLSL
 PELIGIMDTCFCCCLITWLEGHSLAQVFTCLYIHNPDFIEDPAMKAFALGILKICDIA
 REKVNKAAFEEDFQSMTYGFKMANSVTDLRTGMLKDVEDDLQRRVKSTRSQGEE
 RDPEVELEHQQCCLAAPSRVKFTRVLLTLIIAFTKETSAAVEAQKLMVQAADLLSAIH
 TSLHHGIIQAQNGTTKGDPHPIMMGFEPVLNQRLPPTFPFYAKIIRKEEMVNYFSRLID
 RIKTVCENVNLPNLHCILDDFFCEFESEQSPCVLSRLQTTFLVDNKKVFGTHLMQDMV
 KDALRSFVSPVLPSPKCLYNNHQAQKDCIDSFVTHCVRFCSLVQIHGHNRAQRQDKL
 GHILEEFATLQDEAEKVDAAALHTMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGF
 ELELYSMHEYIYIYVYLSEFLYAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKKK
 KKVRLPSREITMSQAYQNMCAGMFKTMVAFMDGKVRKPKFELDSEQVRYEHRFAPFN
 SVMTPPPVHYLQFKEMSDLNKYSPPPQSPELYVAASKHFQQAQKMLENIPNPDREVSR
 ILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSVHKYFPVVKLV"

misc\_feature
 126..680
 /note="Mak10; Region: Mak10 subunit, NatC
 N(alpha)-terminal acetyltransferase. NatC
 N(alpha)-terminal acetyltransferases contains Mak10p,
 Mak31p and Mak3p subunits. All three subunits are
 associated with each other to form the active complex"
 /db\_xref="CDD:pfam04112"

ORIGIN
 Query Match 44.6%; Score 67.4; DB 10; Length 2594;
 Best Local Similarity 87.1%; Pred. No. 1.1e-10;
 Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||



Db	1513	CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACAACATAGTATGA	1454
QY	70	AGCGCGCATCAACCTTCTCTGCCT	94
Db	1453	AGCGCTGCATCAACCTTCTCTGCCT	1429
RESULT 13			
AF272892/c			
LOCUS	AF272892	2543 bp	linear
DEFINITION	Rattus norvegicus corneal wound healing related protein mRNA, complete cds.		
ACCESSION	AF272892		
VERSION	AF272892.1	GI:8926319	
KEYWORDS	Rattus norvegicus (Norway rat)		
SOURCE	Rattus norvegicus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 2543)		
AUTHORS	Yi,X.J., Li,X.F. and Yu,F.S.		
TITLE	A novel epithelial wound-related gene is abundantly expressed in developing rat cornea and skin		
JOURNAL	Curr. Eye Res. 20 (5), 430-440 (2000)		
MEDLINE	20314365		
PUBMED	10855038		
REFERENCE	2 (bases 1 to 2543)		
AUTHORS	Li,X. and Yu,F.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAY-2000) The Schepens Eye Research, 20 Staniford Street, Boston, MA 02114, USA		
FEATURES	Location/Qualifiers		
source	1. .2543		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	64. .2241		
	/codon_start=1		
	/product="corneal wound healing related protein"		
	/protein_id="AAF81791.1"		
	/db_xref="GI:8926320"		
	/translation="MVKAAVDDDDASGWELNVPEKMEKSSVSDITQDFEDACRELKLGELLDKLFGLFEAMSAIEMDPKMDAGMIGNQVNRKVLNFEQAVKDGITIKIDLSLPELIGIMDTFCCLITWLEGHSLAQTFTCLYHNPDFIEDPAMKAFALGILKICDIAREKVNKAAAFEEDFQSMYGFKMANGVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAAFSRVKFTRVLLTVLIAFTKETSAAVEAQKLMVQAADLLSAIHTSLHGIQAQNGTTKGDHPIMMGFEPLVNRLLPPTFPYAKIIKREEMVNYFSRLIDRIKTVCEVVNLPNLHCILDDFFCEFSQSPCVLSRLIQTTFVLVDNKKVFGTHMQDMVKDAPRSFVSPVLSPKCCLYNNHQAKCIDSFYTHCVRPFCSLVQIHGHNRAQRDKLGHILEEFATLQDAEKVDAALHTMLLKQEPQHQHACLGTWVLYHSLRIMIQYLLSGFDLELYSMHEYIYIYWLSEFLYAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNCAGMFKTMVAFDMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQPKEMSDLSKYSPPPPQPPPELYVAASKHFQQAKMILESIPNADREVSRILKVAKPNFVVMKLLAGCHKKESKVPPEPDFSVHKYFPVVKLV"		
ORIGIN			
Query Match	41.9%;	Score 63.2;	DB 10;
Best Local Similarity	74.1%;	Pred. No. 2.5e-09;	Length 2543;
Matches	80;	Conservative 0;	Mismatches 28; Indels 0; Gaps 0;
QY	10	CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCTCTGTTTCAACAGCATGGTGTGA	69
Db	1535	CAAGTTCCTAAACAGGCCAGATGCTGTCGTGAGGCTCTGCTTCAACAACATAGTGTGG	1476
QY	70	AGCGCGCATCAACCTTCTTCGCCATTATAAATAAATATGTCATAAACT	117
Db	1475	AGCGCTGCATCGACCTTCTTCGCCTCATCTTGCAAGGTGGCAAACTCT	1428
RESULT 14			
BC053286/c			
LOCUS	BC053286	2467 bp	mRNA
			linear
			VRT 07-OCT-2003

DEFINITION	Danio rerio cDNA clone MGC:64157 IMAGE:6797204, complete cds.
ACCESSION	BC053286
VERSION	BC053286.1
KEYWORDS	GI:31418975
SOURCE	MGC.
ORGANISM	Danio rerio (zebrafish)
	Danio rerio
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 2467)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 2467)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-re@mail.nih.gov Tissue Procurement: Leonard I. Zon, M.D. cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 117 Row: n Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.
source	Location/Qualifiers
	1. .2467
	/organism="Danio rerio"
	/mol_type="mRNA"
	/db_xref="taxon:7955"
	/clone="MGC:64157 IMAGE:6797204"
	/tissue type="Kidney, zebrafish"
	/clone_lib="NCI CGAP_ZKId1"
	/lab_host="DH10B"
	/note="Vector: PCMV-SPORT6.1"
	58. .2232
	/codon_start=1
	/product="Unknown (protein for MGC:64157)"
	/protein_id="AAH53286.1"



/db\_xref="GI:31418976"  
/translation="MVMKSSVEEEEGWGLGIPKMRNNANWVDVTOEFKGACKELKL  
GELLHDXLFGLEFAMSAIEMDPKMDAGMIGNQVNRKVLNFDQAVKDEAIRVKDLSIP  
ELIGIMDTFCFCLITWLEGHSLAQTVFTCLYVHNPDLIQDPALKAFALGILKICDIIAR  
EKVGNAAVFEEDFQAMTYGFKMANNVTDLRVTGMLKDVEDELQRKVKSTRSQGEQR  
DPEVELDHQOCLAFSRVKFTRLILLSALISFTKKTSAVSEAQKLMSQAADLIPAVHA  
TIQYGIQSONDTTKGDHPINMGFEPLVNQRLLPPTFRYAKIIKREEMVNYFSKLIER  
IKSVCEVINITNLHSILDFECFEFSEQPCLVRSLLQTTFLIDNKKVFGTHLMQDMIK  
DALRCFVSPPLVSSKSNNNHQAQYIDSFVTHCTFRPFCSLIQIHGHNRRQRDKIG  
HILEEFATLQDEAEKVDAALHGLLMKLEPQRQHLACLGTWILYHNLIRIMIQYLLSGFE  
LELYSMHEYIYIYWLSEFLYAWLMSTLSRADSSQMAEERILEEQLKVRSSKSKKKKK  
KARPLSKEITMSQAYQNCAGMYKTMIALDMDRKVRKPFQFELDSEQVRYEHRFAFPNS  
VVTPPPVIYIOFKEMSDLKKYNPPRPSADLYMAASKHFOQAKLLLENVTSPDAEVNRI  
LKVAKPNIVVMKLLAGGHKKETKALPEFDFS AHKYFFIIVKIL"  
139. .693  
/note="Mak10; Region: Mak10 subunit, NatC  
N(alpha)-terminal acetyltransferase. NatC  
N(alpha)-terminal acetyltransferases contains Mak10p,  
Mak31p and Mak3p subunits. All three subunits are  
associated with each other to form the active complex"  
/db\_xref="CDD:pfam04112"

ORIGIN

Query Match 35.1%; Score 53; DB 5; Length 2467;  
Best Local Similarity 76.5%; Pred. No. 4.8e-06;  
Matches 65; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGGCTTTCCTGTTTCAACAGCATGGTGTA 69  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1526 CAGGTGCCACAGACAGGCTAAATGCTGCCGTGCGGCTCCAGCTTCATCAGCAGACCGTGA 1467  
  
QY 70 AGCGCGCATCAACCTTCTCTGCCT 94  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1466 AGAGCAGCGTCACCTTCTCTGCCT 1442

RESULT 15

AC102010/c  
LOCUS AC102010 147876 bp DNA linear HTG 21-MAR-2003  
DEFINITION Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered  
pieces.  
ACCESSION AC102010  
VERSION AC102010.3 GI:29135693  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 147876)  
Birren,B., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP24-531M24  
Unpublished  
2 (bases 1 to 147876)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

TITLE

JOURNAL

REFERENCE

AUTHORS

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 147876)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 21, 2003 this sequence version replaced gi:28412092.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: LI7873  
Center clone name: 531\_M\_24  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 146099 bases at least Q40  
Consensus quality: 146869 bases at least Q30  
Consensus quality: 147004 bases at least Q20  
Insert size: 157000; agarose-fp  
Insert size: 147276; sum-of-contigs  
Quality coverage: 9.9 in Q20 bases; agarose-fp  
Quality coverage: 10.6 in Q20 bases; sum-of-contigs  
-----

TITLE

JOURNAL

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 84958: contig of 84958 bp in length  
\* 84959 85058: gap of 100 bp  
\* 85059 87411: contig of 2353 bp in length  
\* 87412 87511: gap of 100 bp  
\* 87512 89933: contig of 2422 bp in length  
\* 89934 90033: gap of 100 bp  
\* 90034 94076: contig of 4043 bp in length  
\* 94077 94176: gap of 100 bp  
\* 94177 102536: contig of 8360 bp in length  
\* 102537 102636: gap of 100 bp  
\* 102637 118954: contig of 16318 bp in length  
\* 118955 119054: gap of 100 bp

```
FEATURES
source
* 119055 147876: contig of 28822 bp in length.
Location/Qualifiers
1. .147876
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-531M24"
/clone_lib="RPCI-24 Male Mouse BAC"
misc_feature
1. .84958
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
85059. .87411
/note="assembly_fragment"
misc_feature
87512. .89933
/note="assembly_fragment"
misc_feature
90034. .94076
/note="assembly_fragment"
misc_feature
94177. .102536
/note="assembly_fragment"
misc_feature
102637. .118954
/note="assembly_fragment"
misc_feature
119055. .147876
/note="assembly_fragment"
clone_end:T7
vector_side:right"
```

ORIGIN

```
Query Match 35.1%; Score 53; DB 2; Length 147876;
Best Local Similarity 73.1%; Pred.No. 8.1e-06;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 10 CAGGTACTTAAACAGGCCAAATGTTGCTTTGGGGTTCTCTGTTTCAACAGCATGGTGGA 69
Db 72729 CAGATTCTTAAACAGGTTCAGATGTAGTCTCTCAGGCTCCTGTTTCAACACGGAGTGGA 72670

QY 70 AGCGCCGCATCAACCTTCTCTGCCTATTAAAT 102
Db 72669 AGCACTGCATCAAAAGTTGTCTGCAAGTGTA 72637
```

Search completed: August 11, 2004, 01:34:02  
Job time : 605.675 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:23 ; Search time 410.913 Seconds  
(without alignments)  
10973.574 Million cell updates/sec

Title: US-10-001-857-41  
Perfect score: 151  
Sequence: 1 ccgccgggcagggtacctaa.....gcaaattcctcaagaatatg 151

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_estfun:\*  
15: em\_estom:\*  
16: em\_gss\_hum:\*  
17: em\_gss\_inv:\*  
18: em\_gss\_pln:\*  
19: em\_gss\_vrt:\*  
20: em\_gss\_fun:\*  
21: em\_gss\_mam:\*  
22: em\_gss\_mus:\*  
23: em\_gss\_pro:\*  
24: em\_gss\_rod:\*  
25: em\_gss\_phg:\*  
26: em\_gss\_vrl:\*  
27: gb\_gss1:\*  
28: gb\_gss2:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	104.8	69.4	701	29	AG180176	AG180176 Pan trogl
C 2	86.8	57.5	3071	11	AK050805	AK050805 Mus muscu
C 3	75.2	49.8	812	14	CD654321	CD654321 AGENCOURT
C 4	73.8	48.9	472	14	CF762228	CF762228 CES000418

C 5	73.8	48.9	486	12	BI537770	BI537770 427999 MA
C 6	73.8	48.9	569	12	BI682349	BI682349 463408 MA
C 7	73.8	48.9	814	14	CB169246	CB169246 VBB603020
C 8	72.6	48.1	630	28	AZ870091	AZ870091 2M0182N22
C 9	68.4	45.3	974	10	BF179099	BF179099 601808136
C 10	67.4	44.6	409	9	AI020180	AI020180 ub15b06.r
C 11	67.4	44.6	426	10	BB727790	BB727790 BB727790
C 12	67.4	44.6	435	10	BB827250	BB827250 BB827250
C 13	67.4	44.6	446	10	BB832780	BB832780 BB832780
C 14	67.4	44.6	478	10	BE308753	BE308753 601091825
C 15	67.4	44.6	512	14	W54490	W54490 md09d10.r1
C 16	67.4	44.6	514	10	BF465930	BF465930 UI-M-CG0p
C 17	67.4	44.6	528	10	BE692994	BE692994 SAC448 Mo
C 18	67.4	44.6	528	10	BE693019	BE693019 SAC473 Mo
C 19	67.4	44.6	578	14	CA533399	CA533399 C0348H03-
C 20	67.4	44.6	618	14	CF175090	CF175090 B0948F03-
C 21	67.4	44.6	763	13	BU613022	BU613022 UI-M-FR0-
C 22	67.4	44.6	873	9	AU081000	AU081000 AU081000
C 23	67.4	44.6	900	14	CA977646	CA977646 AGENCOURT
C 24	66.4	44.0	595	10	BF730939	BF730939 mab79a08.
C 25	66.4	44.0	600	13	BU920897	BU920897 6071-74 M
C 26	65.8	43.6	434	10	BB811746	BB811746 BB811746
C 27	63.2	41.9	565	12	BM387958	BM387958 UI-R-CN1-
C 28	63.2	41.9	708	14	CF113422	CF113422 Shultzomi
C 29	62	41.1	416	10	BB820704	BB820704 BB820704
C 30	61.6	40.8	437	10	BF522938	BF522938 UI-R-Y0-a
C 31	61.6	40.8	482	14	CA792702	CA792702 AGENCOURT
C 32	61.6	40.8	603	9	AL782358	AL782358 AL782358
C 33	61.6	40.8	868	14	CF238691	CF238691 AGENCOURT
C 34	61.4	40.7	518	13	BY476464	BY476464 BY476464
C 35	57.8	38.3	300	10	BB192400	BB192400 BB192400
C 36	57.8	38.3	784	13	BU260643	BU260643 603503268
C 37	57.8	38.3	833	13	BU464042	BU464042 603773152
C 38	56.4	37.4	399	14	CF727314	CF727314 UI-M-HB0-
C 39	55.6	36.8	236	10	BF172850	BF172850 MYE0095 M
C 40	55	36.4	797	13	BX870686	BX870686 BX870686
C 41	54.6	36.2	303	10	BB273684	BB273684 BB273684
C 42	53.8	35.6	288	10	BB099049	BB099049 BB099049
C 43	53.8	35.6	456	10	BB730820	BB730820 BB730820
C 44	50.4	33.4	302	10	BB314183	BB314183 BB314183
C 45	50.2	33.2	398	10	BG062030	BG062030 L0955F10-

ALIGNMENTS

RESULT 1  
AG180176 701 bp DNA linear GSS 09-JAN-2002  
LOCUS Pan troglodytes DNA, clone: RP43-052K19.T7, genomic survey  
DEFINITION sequence.  
ACCESSION AG180176  
VERSION AG180176.1 GI:16709856  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE 1  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of Library RPCI-43  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 701)  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC

end was generated during the RED process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
Location/Qualifiers  
1. .701  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-052K19.T7"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Query Match 69.4%; Score 104.8; DB 29; Length 701;  
Best Local Similarity 98.1%; Pred. No. 6.6e-23;  
Matches 106; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 69  
|||||  
Db 168 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 227  
|||||  
QY 70 AGCGCCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCAATAACT 117  
|||||  
Db 228 AGCGCTGCATCAACCTTGTCTGCCTATTAAAAATAAAATGTCAATAACT 275  
|||||

RESULT 2

AK050805/c

LOCUS

AK050805 3071 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 9 days embryo whole body cDNA, RIKEN full-length  
enriched library, clone:D030020M24 product:CORNEAL WOUND HEALING  
RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.

ACCESSION

AK050805.1 GI:26094130

VERSION

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci,P. and Hayashizaki,Y.

AUTHORS

Carninci,P. and Hayashizaki,Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

AUTHORS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159  
3  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL

20530913

MEDLINE

11076861

PUBMED

11076861

REFERENCE

4

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5

The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3071)

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.

URL:http://genome.gsc.riken.go.jp/

URL:http://fantom.gsc.riken.go.jp/.

Location/Qualifiers

1. .3071

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:D030020M24"

/db\_xref="MGI:2418664"

/db\_xref="taxon:10090"

/clone="D030020M24"

/tissue\_type="whole body"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev stage="9 days embryo"

1. .3071

/note="CORNEAL WOUND HEALING RELATED PROTEIN homolog

[Rattus norvegicus] (SPTR|Q9JI01, evidence: FASTY,

96.2%ID, 43.5%length, match=948)"

ORIGIN

Query Match 57.5%; Score 86.8; DB 11; Length 3071;

Best Local Similarity 81.0%; Pred. No. 5.8e-17;

Matches 115; Conservative 0; Mismatches 22; Indels 5; Gaps 1;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 69

|||||

Db 939 CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACAACATAGTAGA 880

|||||

QY 70 AGCGCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCAATAACTCATCTGCAAGG 129

|||||

Db 879 AGCGCTGCATCAACCTTCTCTGCCTGTCAAAACAATTGTT-----ACTCATCTTGCAAGG 825

|||||

QY 130 TGGCAAATTCCTCAAGAATATG 151

|||||

Db 824 TAGCAAACCTCCTCAAGAATGTG 803

|||||



RESULT 3  
CD654321/c  
LOCUS  
DEFINITION  
AGENCOURT\_14553680 NIA Human H1 Embryonic Stem Cell cDNA Library  
(Long) Homo sapiens cDNA clone IMAGE:30428535 5', mRNA sequence.  
CD654321  
CD654321.1 GI:31893486  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 812)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Irene Ginis and Mahendra Rao, NIA  
cDNA Library Preparation: Yulan Piao and Minoru Ko  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC c lone distribution information  
can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM517 row: 1 column: 16  
High quality sequence stop: 683.  
FEATURES  
source  
1..812  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30428535"  
/tissue type="Embryonic Stem cells"  
/cell\_line="WA01"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIA Human H1 Embryonic Stem Cell cDNA Library  
(Long)"  
/note="Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI;  
This is a long-transcript enriched cDNA library (Genome  
Res. 11: 1553-1558 (2001). [PMID: 11544199]) from WA01  
cell line. Undifferentiated human ES cell line WA01/H1  
was obtained from WiCell Research Institute, Inc.,  
Madison, WI, cultured according to their instructions, on  
MEF feeders. They formed round colonies with defined edges  
and were positive for alkaline phosphatase, SSEA-4, OCT3,  
OCT4, REX1, UTF, TERT, SOX2, CX43 and CX45. They are  
negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1,  
TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days  
after plating), the ES cells from 4 X 6cm dishes were  
treated with 1 mg/ml collagenase, type IV  
(Invitrogen/GIBCO) for 5-10 min and gently scraped off  
with 5 ml pipette. RNA was purified with TRIzol Reagent  
from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558  
(2001). [PMID:11544199]) Double-stranded cDNAs were  
synthesized with an Oligo(dt) primer [Invitrogen:  
5'-pGACTAGTCTAGATCGGAGCGCGCCCTTTT-TTTT-3'] from  
3.4g of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to Lone-linker LL-Sal4, purified by phenol/chloroform  
extraction, and separated from free linkers by  
Centricon-100 column. Then, the cDNAs were amplified by  
long-range high fidelity PCR using Ex Taq polymerase  
(Takara) with a primer Sal4-S for 25 cycles. The products  
were purified by phenol/chloroform extraction and  
Centricon-100 column. The cDNAs were digested with SalI  
and NotI enzymes and cloned into SalI/NotI site of  
pCMV-SPORT6 plasmid vector. The average insert size is  
about 3.6kb."

ORIGIN  
Query Match 49.8%; Score 75.2; DB 14; Length 812;  
Best Local Similarity 95.1%; Pred. No. 2.5e-13;  
Matches 77; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 37 CTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTAT 96  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 730 CCTTTGGGTTTCNTGTTTCAACAGCATGGTGTGAAGCGCTGCATCAACCTTCTCTGCCTAT 671  
QY 97 TAAATAAAATGTCATAAACT 117  
||| ||||| ||||| ||||| |||||  
Db 670 TAAATAAAATGTCATAAACT 650  
RESULT 4  
CF762228/c  
LOCUS  
DEFINITION  
CF762228 Bos taurus skin cDNA library Bos taurus linear EST 17-OCT-2003  
CCL000418 5', mRNA sequence.  
CF762228  
CF762228.1 GI:37711446  
EST.  
KEYWORDS  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 472)  
Wang, Y.H., McWilliam, S. and Lehnert, S.  
Transcription profiling of cattle skin  
Unpublished (2003)  
JOURNAL  
COMMENT  
Contact: Dr Yonghong Wang  
Functional Genomics Lab  
CSIRO Livestock Industries  
Level 5, Queensland Biosciences Precinct, University of Queensland,  
306 Carmody Road St. Lucia QLD Australia  
Tel: 07 3214 2445  
Fax: 07 3214 2685  
Email: [Yonghong.Wang@csiro.au](mailto:Yonghong.Wang@csiro.au)  
Plate: 14 row: D column: 09.  
FEATURES  
source  
1..472  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Hereford Shorthorn"  
/db\_xref="taxon:9913"  
/clone="CCL000418"  
/sex="female"  
/tissue\_type="pooled"  
/dev\_stage="Adult"  
/lab\_host="XL1-BlueMRF' strain"  
/clone\_lib="Bos taurus skin cDNA library"  
/note="Organ: skin; Vector: Uni-ZAPXR; Site\_1: EcoRI;  
Site\_2: Xho I; Library made from pooled skin of adult  
female Hereford-Shorthorn."  
ORIGIN  
Query Match 48.9%; Score 73.8; DB 14; Length 472;  
Best Local Similarity 91.8%; Pred. No. 6.3e-13;  
Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGA 69  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 201 CAGGTACCTAAACAGGCCAAATGTCCTGTGGGGCTCCTGTTTCAACAGCATAGTATGA 142  
QY 70 AGCGCCGCATCAACCTTCTCTGCCT 94  
||| ||| ||||| ||||| |||||  
Db 141 AGGGCTGCATCAACCTTCTCTGCCT 117  
RESULT 5  
BI537770/c

LOCUS BI537770 486 bp mRNA linear EST 30-AUG-2001  
DEFINITION 427999 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BI537770  
VERSION BI537770.1 GI:15378880  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
PUBMED 11282978  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTCCCAGTCACGACG  
Plate: 129 row: J column: 14  
Seq primer: ATTTAGGTGACACTATAG.  
FEATURES  
source Location/Qualifiers  
1..486  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 4BOV"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from day 20 and day 40 embryos."  
ORIGIN  
Query Match 48.9%; Score 73.8; DB 12; Length 486;  
Best Local Similarity 91.8%; Pred. No. 6.4e-13;  
Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTA 69  
|||||  
Db 278 CAGGTACCTAAACAGGCCAAATGTCCTGTGGGGCTCCTGTTTCAACAGCATAGTATGA 219  
QY 70 AGCGCCGCATCAACCTTCTCTGCCT 94  
|||  
Db 218 AGGCTGCATCAACCTTCTCTGCCT 194  
RESULT 6  
BI682349/c  
LOCUS BI682349 569 bp mRNA linear EST 17-SEP-2001  
DEFINITION 463408 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BI682349  
VERSION BI682349.1 GI:15635283  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Bovinae; Bos.  
LOCUS BI537770 486 bp mRNA linear EST 30-AUG-2001  
DEFINITION 427999 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BI537770  
VERSION BI537770.1 GI:15378880  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 569)  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
PUBMED 11282978  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTCCCAGTCACGACG  
Plate: 132 row: B column: 15  
Seq primer: ATTTAGGTGACACTATAG.  
FEATURES  
source Location/Qualifiers  
1..569  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 1BOV"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."  
ORIGIN  
Query Match 48.9%; Score 73.8; DB 12; Length 569;  
Best Local Similarity 91.8%; Pred. No. 6.6e-13;  
Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTA 69  
|||||  
Db 411 CAGGTACCTAAACAGGCCAAATGTCCTGTGGGGCTCCTGTTTCAACAGCATATGA 352  
QY 70 AGCGCCGCATCAACCTTCTCTGCCT 94  
|||  
Db 351 AGGCTGCATCAACCTTCTCTGCCT 327  
RESULT 7  
CB169246  
LOCUS CB169246 814 bp mRNA linear EST 30-JAN-2003  
DEFINITION VBB603020215.R1 CSEQFXN41 testes Bos taurus cDNA, mRNA sequence.  
ACCESSION CB169246  
VERSION CB169246.1 GI:28155372  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 814)  
AUTHORS Adelson,D.L. and Gill,C.A.  
TITLE Bovine ESTs (Adelson and Gill)  
JOURNAL Unpublished (2003)  
COMMENT Contact: David L. Adelson  
Animal Breeding and Genetics  
Texas A&M University  
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471, USA  
Tel: 9798452616

Fax: 9798456970  
Email: david.adelson@tamu.edu.

FEATURES  
source  
1. .814  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="testes"  
/clone\_lib="CSEQFXN41 testes"  
/note="Organ: testes; Vector: pBluescript SK+; Site\_1:  
NotI; Site 2: EcoRI; sequence 5' of the insert  
(5'-NNN. . .NNNinsert)  
GCGAATTGGAGCTCCACCCGGGTGGCGGCGGCTCGAG. Sequence 3' of  
the inserts (AAGAAATCGATATCAAGCTTATCGATACCGTCGACCTCGAG.  
normalized Rd 1 library, sequenced 3' with M13R primer."

ORIGIN  
Query Match 48.9%; Score 73.8; DB 14; Length 814;  
Best Local Similarity 91.8%; Pred. No. 7.1e-13;  
Matches 78; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69  
|||||  
Db 134 CAGGTACCTAAACAGGCCAAATGCTGCCTGTGGGCTCCTGTTTCAACAGCATAGTATGA 193  
|||||

QY 70 AGCGCGCATCAACCTTCTCTGCCT 94  
|||||  
Db 194 AGGCTGCATCAACCTTCTCTGCCT 218  
|||||

RESULT 8  
AZ870091 630 bp DNA linear GSS 21-FEB-2001  
LOCUS 2M0182N22F Mouse 10kb plasmid UUGC1m library Mus musculus genomic  
DEFINITION clone UUGC2M0182N22 F, genomic survey sequence.  
ACCESSION AZ870091  
VERSION AZ870091.1 GI:13075085  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 630)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D.,Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0182 row: N column: 22  
Seq primer: CGTTGTAAACGACGGCCAGT  
Class: plasmid ends  
High quality sequence stop: 630.  
Location/Qualifiers  
1. .630  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0182N22"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1m library"

FEATURES  
source  
1. .630  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0182N22"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1m library"

/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

ORIGIN  
Query Match 48.1%; Score 72.6; DB 28; Length 630;  
Best Local Similarity 85.3%; Pred. No. 1.6e-12;  
Matches 81; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69  
|||||  
Db 380 CAGGTTCTCTAAACAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACACATAGTATGA 439  
|||||

QY 70 AGCGCCGCATCAACCTTCTCTGCCTATTAAATAA 104  
|||||  
Db 440 AGCGCTGCATGAACCTTCTCTGCCTGTCAAAACAA 474  
|||||

RESULT 9  
BF179099 974 bp mRNA linear EST 31-OCT-2000  
LOCUS 601808136F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4038945 5',  
DEFINITION mRNA sequence.  
ACCESSION BF179099  
VERSION BF179099.1 GI:11057241  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 974)  
NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLAM9318 row: g column: 10  
High quality sequence stop: 662.  
Location/Qualifiers  
1. .974  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4038945"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;









Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source  
Location/Qualifiers  
1..446  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="G930023L04"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"

ORIGIN

Query Match 44.6%; Score 67.4; DB 10; Length 446;  
Best Local Similarity 87.1%; Pred. No. 7.3e-11;  
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69  
Db 232 CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACACATAGTATGA 173  
QY 70 AGCGCCGCATCAACCTTCTCTGCCT 94  
Db 172 AGCGCTGCATCAACCTTCTCTGCCT 148

RESULT 14  
BE308753/c  
LOCUS  
DEFINITION  
601091825F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3486246 5',  
mRNA sequence.  
ACCESSION  
BE308753  
VERSION  
BE308753.1 GI:9165894  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 478)  
REFERENCE  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8522 row: b column: 07  
High quality sequence stop: 478.

FEATURES  
source  
Location/Qualifiers  
1..478  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3486246"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

ORIGIN

Query Match 44.6%; Score 67.4; DB 14; Length 512;  
Best Local Similarity 87.1%; Pred. No. 7.6e-11;

Query Match 44.6%; Score 67.4; DB 10; Length 478;  
Best Local Similarity 87.1%; Pred. No. 7.4e-11;  
Matches 74; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69  
Db 315 CAGGTTCTTAAGCAGGCCAGATGTTGTCTCTGAGGCTCCTGTTTCAACACATAGTATGA 256  
QY 70 AGCGCCGCATCAACCTTCTCTGCCT 94  
Db 255 AGCGCTGCATCAACCTTCTCTGCCT 231

RESULT 15

W54490/c  
LOCUS  
DEFINITION  
W54490 md09d10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
clone IMAGE:367891 5', mRNA sequence.  
ACCESSION  
W54490  
VERSION  
W54490.1 GI:1355550  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 512)  
REFERENCE  
AUTHORS  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE  
The WashU-HHMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:229323  
Seq primer: ETprimer  
High quality sequence stop: 344.  
Location/Qualifiers  
1..512  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:367891"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NbME13.5 14.5"  
/note="Vector: pF7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGGCCGGAATTTTCTTTTCTTTTCTTTTCTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pF7T3 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."

FEATURES

source  
Location/Qualifiers  
1..512  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:367891"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NbME13.5 14.5"  
/note="Vector: pF7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGGCCGGAATTTTCTTTTCTTTTCTTTTCTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pF7T3 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."

ORIGIN

Query Match 44.6%; Score 67.4; DB 14; Length 512;  
Best Local Similarity 87.1%; Pred. No. 7.6e-11;

	Matches	74;	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
Qy	10	CAGGTACCTAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA	69							
Db	404	CAGGTTCCCTAAGCAGGCCAGATGTTGTTCTCTGAGGCTCCTGTTTCAACACATAGTATGA	345							
Qy	70	AGCGCCGCATCAACCTTCTCTGCCT	94							
Db	344	AGCGCTGCATCAACCTTCTCTGCCT	320							

Search completed: August 11, 2004, 04:18:35  
Job time : 415.913 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:17 ; Search time 60.5488 Seconds  
(without alignments)  
10594.393 Million cell updates/sec

Title: US-10-001-857-41  
Perfect score: 151  
Sequence: 1 ccgcccgggcagggtacctaa.....gcaaatctcctaagaatatg 151

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	%			Query Match	Length	DB ID	Description
	Score	Match	Length				
1	151	100.0	151	6	ABQ75302		Abq75302 Human lun
C 2	140.4	93.0	3096	6	ABQ75303		Abq75303 Human lun
C 3	84.8	56.2	2488	6	AAL49929		Aal49929 Human mol
4	83.8	55.5	303	5	ABV06162		Abv06162 Human pro
5	83.8	55.5	439	5	ABV45182		Abv45182 Human pro
6	83.8	55.5	439	5	ABV36125		Abv36125 Human pro
7	83.8	55.5	815	6	ABQ75372		Abq75372 Human lun
8	83.2	55.1	2668	5	ABV29823		Abv29823 Human pro
9	83.2	55.1	2668	5	ABV23940		Abv23940 Human pro
C 10	83.2	55.1	3029	9	ADC30041		Adc30041 Human nov
11	82.8	54.8	356	5	ABV15331		Abv15331 Human pro
C 12	63.2	41.9	2543	9	ADB53632		Adb53632 Primary r
C 13	33.4	22.1	3215	6	ABL55204		Ab155204 HBV subty
14	32.2	21.3	129021	3	AAF22296		Aaf22296 BAC conta
C 15	31.8	21.1	560	7	ABT23617		Abt23617 Stabilisi
C 16	31.8	21.1	3214	1	AAN60714		Aan60714 Sequence
C 17	31.8	21.1	160755	4	AAH88704		Aah88704 Human DNA
C 18	31	20.5	1445	2	AAV82684		Aav82684 Fulminant
C 19	31	20.5	1500	2	AAV82693		Aav82693 Fulminant
C 20	31	20.5	1644	7	ACA35581		Aca35581 Prokaryot
21	30.8	20.4	362	5	ABV32031		Abv32031 Human pro
22	30.8	20.4	487	5	ABV40970		Abv40970 Human pro
23	30.6	20.3	322101	9	AAD58431		Aad58431 Human PAO

C 24	30.2	20.0	462	2	AAT05544	Aat05544 Human hep
C 25	30.2	20.0	560	7	ABT23618	Abt23618 Stabilisi
C 26	30.2	20.0	3214	2	AAQ05377	Aaq05377 Fragment
C 27	30.2	20.0	3835	2	AAQ05378	Aaq05378 Sequence
C 28	30.2	20.0	4421	2	AAT58319	Aat58319 DNA seque
C 29	30	19.9	13273	4	AAS36849	Aas36849 Human car
C 30	30	19.9	13273	9	ADE47543	Ade47543 Human car
C 31	30	19.9	35057	7	AAL51501	Aal51501 Human pan
C 32	30	19.9	300000	9	ADE86352	Ade86352 Human ptp
C 33	29.6	19.6	576	7	ABZ39816	Abz39816 N. gonorr
C 34	29.6	19.6	576	7	ACA41490	Aca41490 Prokaryot
C 35	29.6	19.6	1035	3	AAZ54048	Aaz54048 Neisseria
C 36	29.4	19.5	197	6	AAS16091	Aas16091 Hepatitis
C 37	29.4	19.5	250	6	ABK29867	Abk29867 Wild type
C 38	29.4	19.5	1445	2	AAV82692	Aav82692 Fulminant
C 39	29.4	19.5	1445	2	AAV82685	Aav82685 Fulminant
C 40	29.4	19.5	1445	2	AAV82690	Aav82690 Fulminant
C 41	29.4	19.5	1500	2	AAV82694	Aav82694 Fulminant
C 42	29.4	19.5	1500	2	AAV82697	Aav82697 Fulminant
C 43	29.4	19.5	1500	2	AAV82686	Aav82686 Fulminant
C 44	29.4	19.5	2342	1	AAN93072	Aan93072 Sequence
C 45	29.4	19.5	3182	6	AAD31765	Aad31765 Hepatitis

ALIGNMENTS

RESULT 1  
ABQ75302  
ID ABQ75302 standard; cDNA; 151 BP.  
XX  
AC ABQ75302;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Human lung specific nucleic acid sequence SEQ ID NO:41.

KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma; gene; chromosome 9; ss.

OS Homo sapiens.

PN WO200264788-A2.

PD 22-AUG-2002.

PF 20-NOV-2001; 2001WO-US045080.

PR 20-NOV-2000; 2000US-0252054P.

XX (DIAD-) DIADEXUS INC.

PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;

DR WPI; 2002-657601/70.

XX New lung specific nucleic acid useful in gene therapy or as vaccines for treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung diseases, as well as for diagnosing, monitoring or staging these diseases.

PS Claim 1; Page 168; 282pp; English.

XX The present invention describes an isolated lung specific nucleic acid (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965; (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp), given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b); or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific protein (LSP) sequences have cytostatic activity and can be used in gene therapy and vaccines. LSNAs and LSPs are useful for diagnosing and monitoring the presence and metastases of lung cancer in a patient. An

CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against the  
CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs  
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,  
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-  
CC cancerous disease states in lung

XX  
SQ Sequence 151 BP; 42 A; 39 C; 32 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 151; DB 6; Length 151;  
Best Local Similarity 100.0%; Pred. No. 3.6e-43;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 60

Db 1 CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 60

QY 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATATAAATGTATATAAATCAT 120

Db 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATATAAATGTATATAAATCAT 120

QY 121 CCTGCAAGGTGGCAAAATTCCTCAAGAAATATG 151

Db 121 CCTGCAAGGTGGCAAAATTCCTCAAGAAATATG 151

RESULT 2

ABQ75303/c  
ID ABQ75303 standard; cDNA; 3096 BP.

XX  
AC ABQ75303;

XX  
DT 05-NOV-2002 (first entry)

XX  
DE Human lung specific nucleic acid sequence SEQ ID NO:42.

XX  
KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma; gene; chromosome 9; ss.

XX  
OS Homo sapiens.

XX  
PN WO200264788-A2.

XX  
PD 22-AUG-2002.

XX  
PF 20-NOV-2001; 2001WO-US045080.

XX  
PR 20-NOV-2000; 2000US-0252054P.

XX  
PA (DIAD-) DIADEXUS INC.

XX  
PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;

XX  
DR WPI; 2002-657601/70.

XX  
PT New lung specific nucleic acid useful in gene therapy or as vaccines for  
PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung  
PT diseases, as well as for diagnosing, monitoring or staging these  
PT diseases.

XX  
PS Claim 1; Page 168-169; 282pp; English.

XX  
CC The present invention describes an isolated lung specific nucleic acid  
CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid  
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;  
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),  
CC given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b);  
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific  
CC protein (LSP) sequences have cytostatic activity and can be used in gene  
CC therapy and vaccines. LSNAs and LSPs are useful for diagnosing and  
CC monitoring the presence and metastases of lung cancer in a patient. An

CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against the  
CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs  
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,  
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-  
CC cancerous disease states in lung

XX  
SQ Sequence 3096 BP; 914 A; 606 C; 749 G; 827 T; 0 U; 0 Other;

Query Match 93.0%; Score 140.4; DB 6; Length 3096;  
Best Local Similarity 99.3%; Pred. No. 6.9e-39;  
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 69

Db 2052 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGA 1993

QY 70 AGCGCGGCATCAACCTTCTCTGCCTATTAAATAAAATGTATATAAATCATCTGCAAGG 129

Db 1992 AGCGCTGCATCAACCTTCTCTGCCTATTAAATAAAATGTATATAAATCATCTGCAAGG 1933

QY 130 TGGCAAAATTCCTCAAGAATATG 151

Db 1932 TGGCAAAATTCCTCAAGAATATG 1911

RESULT 3

AAL49929/c

ID AAL49929 standard; cDNA; 2488 BP.

XX  
AC AAL49929;

XX  
DT 10-DEC-2002 (first entry)

XX  
DE Human molecule for disease detection and treatment coding sequence #3.

XX  
KW Human; molecule for disease detection and treatment; MDDT; gene therapy;  
KW cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;  
KW antiinflammatory; antiasthmatic; cerebroprotective; nootropic;  
KW neuroprotective; antiparkinsonian; cardiant; antianginal; gene; ss.

XX  
OS Homo sapiens.

XX  
PN WO200270709-A2.

XX  
PD 12-SEP-2002.

XX  
PF 08-FEB-2002; 2002WO-US003709.

XX  
PR 09-FEB-2001; 2001US-0268117P.

XX  
PR 15-FEB-2001; 2001US-0269618P.

XX  
PR 23-FEB-2001; 2001US-0271118P.

XX  
PR 07-MAR-2001; 2001US-0274486P.

XX  
PR 09-MAR-2001; 2001US-0274436P.

XX  
PR 28-NOV-2001; 2001US-0334229P.

XX  
PR 01-FEB-2002; 2002US-0353284P.

XX  
PA (INCY-) INCYTE GENOMICS INC.

XX  
PI Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;

PI Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;

PI Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;

PI Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;

PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;

XX  
DR WPI; 2002-713453/77.

XX  
DR P-PSDB; AAO19400.

XX  
PT New human molecules for disease detection and treatment (MDDT), useful  
PT for diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant MDDT expression, e.g. cancer, AIDS, asthma, diabetes,  
PT hepatitis.

XX Claim 11; Page 159-160; 177pp; English.

PS The present invention relates to human proteins and coding sequences of

XX molecules for disease detection and treatment MDDT. The sequences can be

CC used in the treatment of diseases associated with the decreased

CC expression or overexpression of MDDT, such as cell proliferative (cancer,

CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,

CC allergies, Addison's disease, asthma), developmental (dwarfism, renal

CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,

CC epilepsy) and cardiovascular (congestive heart failure, myocardial

CC infarction, angina pectoris) disorders. The present sequence is a coding

CC sequence of the invention

XX Sequence 2488 BP; 768 A; 461 C; 567 G; 692 T; 0 U; 0 Other;

SQ

Query Match 56.2%; Score 84.8; DB 6; Length 2488;

Best Local Similarity 88.5%; Pred. No. 2.8e-19;

Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTCAACAGCATGGTGTGA 69

Db

1554 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTCAACAGCATGGTGTGA 1495

QY 70 AGCGCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCATA 113

Db

1494 AGCGCTGCATCAACCTTCTCTGCCTCATCTGCTGCAAGGTGGCAA 1451

RESULT 4

ABV06162

ID ABV06162 standard; cDNA; 303 BP.

XX

AC ABV06162;

XX

DT 13-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 6153.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynanamic marker;

KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.

XX

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI; 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 1018; 11750pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)

CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynanamic or pharmacogenomic marker

XX

SQ Sequence 303 BP; 81 A; 72 C; 77 G; 73 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 303;

Best Local Similarity 88.3%; Pred. No. 2.7e-19;

Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTCAACAGCATGGTGTGAA 70

Db

22 AGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTCAACAGCATGGTGTGAA 81

QY 71 GCGCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCATA 113

Db

82 GCGCTGCATCAACCTTCTCTGCCTCATCTGCTGCAAGGTGGCAA 124

RESULT 5

ABV45182

ID ABV45182 standard; cDNA; 439 BP.

XX

AC ABV45182;

XX

DT 16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 45173.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynanamic marker;

KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.

XX

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI; 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 8947; 11750pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)



CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 439;  
Best Local Similarity 88.3%; Pred. No. 3.2e-19;  
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGGCCAAATGTTGCCTTGGGGTTCCTGTTTCAACAGCATGGTGAA 70  
|||||  
Db 100 AGGTACCTAAACAGGCCAAATGTTGCCTTGGGGTTCCTGTTTCAACAGCATGGTGAA 159  
|||||

QY 71 GCGCCGATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
|||||  
Db 160 GCGTGATCAACCTTCTCTGCCTCATCTGCAAGGTGGCAA 202  
|||||

RESULT 6  
ABV36125  
ID ABV36125 standard; cDNA; 439 BP.  
XX  
AC ABV36125;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 36116.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US005171.  
XX  
PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 7483; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;

Query Match 55.5%; Score 83.8; DB 5; Length 439;  
Best Local Similarity 88.3%; Pred. No. 3.2e-19;  
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 AGGTACCTAAACAGGCCAAATGTTGCCTTGGGGTTCCTGTTTCAACAGCATGGTGAA 70  
|||||  
Db 100 AGGTACCTAAACAGGCCAAATGTTGCCTTGGGGTTCCTGTTTCAACAGCATGGTGAA 159  
|||||

QY 71 GCGCCGATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
|||||  
Db 160 GCGTGATCAACCTTCTCTGCCTCATCTGCAAGGTGGCAA 202  
|||||

RESULT 7  
ABQ75372  
ID ABQ75372 standard; cDNA; 815 BP.  
XX  
AC ABQ75372;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Human lung specific nucleic acid sequence SEQ ID NO:111.  
XX  
KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma; gene; chromosome 9; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200264788-A2.  
XX  
PD 22-AUG-2002.  
XX  
PF 20-NOV-2001; 2001WO-US045080.  
XX  
PR 20-NOV-2000; 2000US-0252054P.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;  
XX  
DR WPI; 2002-657601/70.  
XX  
PT New lung specific nucleic acid useful in gene therapy or as vaccines for  
PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung  
PT diseases, as well as for diagnosing, monitoring or staging these  
PT diseases.  
XX  
PS Claim 1; Page 223; 282pp; English.  
XX  
CC The present invention describes an isolated lung specific nucleic acid  
CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid  
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;  
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),  
CC given in ABQ75262 to ABQ75376; (c) selectively hybridizes to (a) or (b);  
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific  
CC protein (LSP) sequences have cytostatic activity and can be used in gene  
CC therapy and vaccines. LSNA and LSPs are useful for diagnosing and  
CC monitoring the presence and metastases of lung cancer in a patient. An  
CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against the  
CC lung cancer cell expressing the LSNA or LSPs. In particular, these LSNA  
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,  
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-  
CC cancerous disease states in lung  
XX  
SQ Sequence 815 BP; 310 A; 116 C; 207 G; 82 T; 0 U; 100 Other;

Query Match 55.5%; Score 83.8; DB 6; Length 815;  
Best Local Similarity 88.3%; Pred. No. 4.1e-19;  
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;



RESULT 10  
ADC30041/c  
ID ADC30041 standard; cDNA; 3029 BP.  
XX AC ADC30041;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human novel cDNA sequence, SEQ ID NO:123.  
XX KW Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 9; gene; ss.  
XX OS Homo sapiens.  
XX PN WO2003029271-A2.  
XX PD 10-APR-2003.  
XX PF 24-SEP-2002; 2002WO-US030474.  
XX PR 24-SEP-2001; 2001US-0324631P.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
DR WPI; 2003-371981/35.  
DR P-PSDB; ADC31012.  
XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
XX PS Claim 1; SEQ ID NO 123; 1185pp; English.  
XX CC The invention relates to 971 novel human cDNA sequences (ADC29919-  
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the  
CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition; kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC contig sequences corresponding to the cDNA sequences of the invention  
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
CC -ADC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically

CC claimed human cDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 3029 BP; 895 A; 593 C; 739 G; 802 T; 0 U; 0 Other;  
Query Match 55.1%; Score 83.2; DB 9; Length 3029;  
Best Local Similarity 87.5%; Pred.No. 1.1e-18;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 69  
Db 1983 CAGGTACCTAAACAGGCCAAATGTTGCTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 1924  
QY 70 AGCGCCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
Db 1923 AGCGCTGCATCAACCTTCTCTGCCTATCCTGCAAGGTGGCAA 1880  
RESULT 11  
ABV15331  
ID ABV15331 standard; cDNA; 356 BP.  
XX AC ABV15331;  
XX DT 13-SEP-2002 (first entry)  
XX DE Human prostate expression marker cDNA 15322.  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX PD 23-AUG-2001.  
XX PF 20-FEB-2001; 2001WO-US005171.  
XX PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX PI Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX PS Claim 1; Page 2572; 11750pp; English.  
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX



SQ Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 U; 0 Other;

Query Match 54.8%; Score 82.8; DB 5; Length 356;  
Best Local Similarity 88.2%; Pred. No. 6.6e-19;  
Matches 90; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 12 GGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCCCTGTTCAACAGCATGGTGTGAAG 71  
|||||  
Db 1 GGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCCCTGTTCAACAGCATGGTGTGAAG 60  
|||||

QY 72 CGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
|||||  
Db 61 CGCTGCATCAACCTTCTCTGCCTCATCCTGCAAGGTGGCAA 102  
|||||

RESULT 12  
ADB53632/c

ID ADB53632 standard; DNA; 2543 BP.

AC ADB53632;

XX

DT 04-DEC-2003 (first entry)

XX

DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4174.

XX

KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;

KW toxicity marker; toxicity progression; drug screening;

KW primary rat hepatocyte toxicity modelling; gene; ds.

XX

OS Rattus norvegicus.

XX

PN WO2003065993-A2.

XX

PD 14-AUG-2003.

XX

PF 04-FEB-2003; 2003WO-US003482.

XX

PR 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 10-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-MAY-2002; 2002US-0378665P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX

PA (GENE-) GENE LOGIC INC.

XX

PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

PI Blashoff M;

XX

DR WPI; 2003-731472/69.

XX

PT Determining if a compound induces a toxic effect on a tissue or cell, for

PT identifying hepatotoxic compounds, comprises comparing a gene expression

PT profile of a tissue or cell sample to a database of Tox mean and non-Tox

PT mean values.

XX

PS Claim 44; SEQ ID NO 4174; 874pp; English.

XX

CC The present invention describes a method for determining whether a

CC compound induces a toxic effect on a tissue or cell. The method comprises

CC preparing a gene expression profile of a tissue or cell sample exposed to

CC the compound, and comparing the gene expression profile to a database

CC comprising data or information on the Tox mean and non-Tox mean value.

CC The method is useful for predicting or identifying at least one toxic

CC effect, particularly hepatotoxicity, of a test or unknown compound. The

CC genes listed in the specification are useful as diagnostic or toxicity

CC markers for the prediction or identification of the physiological state

CC of tissue or cell sample that has been exposed to a compound, or to

CC identify or predict the toxic effects of a compound or an agent. These

CC may also be used as markers for monitoring toxicity progression or for

CC drug screening. The present sequence represents a primary rat hepatocyte

CC toxicity modelling related gene sequence from the present invention.

XX

SQ Sequence 2543 BP; 725 A; 554 C; 627 G; 637 T; 0 U; 0 Other;

Query Match 41.9%; Score 63.2; DB 9; Length 2543;  
Best Local Similarity 74.1%; Pred. No. 1.2e-11;  
Matches 80; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCCCTGTTCAACAGCATGGTGTGA 69  
|||||

Db 1535 CAAGTCTCTAAACAGGCCAGATGCTGCTGAGGCTCCTGCTTCAACAACATAGTGTGG 1476  
|||||

QY 70 AGCGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATAAACT 117  
|||||

Db 1475 AGCGTGCATCGACCTTCTCGCTCATCTTGCAAGGTGGCAAACTCT 1428  
|||||

RESULT 13  
ABL55204/c

ID ABL55204 standard; DNA; 3215 BP.

XX

AC ABL55204;

XX

DT 14-JUN-2002 (first entry)

XX

DE HBV subtype adr surface antigen (HBsAg)/core antigen encoding DNA.

XX

KW HBV; subtype adr; surface antigen; HBsAg; core antigen; detection;

KW hepatitis B; gene; ds.

XX

OS Hepatitis B virus.

XX

FH Key Location/Qualifiers

FT primer\_bind complement(15. .34)

FT /\*tag= a

FT /note= "Primer S-N1/I (ABL55201)"

FT CDS 29. .709

FT /\*tag= b

FT /product= "Surface antigen (HBsAg; AAM49180)"

FT primer\_bind complement(60. .79)

FT /\*tag= c

FT /note= "Primer S-02/I (ABL55207)"

FT primer\_bind complement(70. .89)

FT /\*tag= d

FT /note= "Primer S-03/I (ABL55208)"

FT misc\_binding complement(150. .290)

FT /\*tag= e

FT /bound moiety= "HBV capture probe (ABL55199)"

FT misc\_binding 221. .250

FT /\*tag= f

FT /bound moiety= "AsCap7-Inosine probe (ABL55205)"

FT primer\_bind 698. .723

FT /\*tag= g

FT /note= "Primer AS-2/I (ABL55203)"

FT primer\_bind 710. .733

FT /\*tag= h

FT /note= "Primer AS-3/I (ABL55202)"

FT primer\_bind 721. .745

FT /\*tag= i

FT /note= "Primer AS-4/I (ABL55209)"

FT CDS 1775. .2336

FT /\*tag= j

FT /product= "Core antigen (AAM49181)"



```
FT misc_feature 2780..3215
FT /*tag= k
FT /note= "Corresponds to bases 1-436 of ABL55200"
XX
PN JP2001352989-A.
XX
XX 25-DEC-2001.
XX
XX 14-JUN-2000; 2000JP-00178470.
XX
XX 14-JUN-2000; 2000JP-00178470.
XX
XX (FJRE ) FUJIREBIO KK.
XX
DR WPI; 2002-191858/25.
DR P-PSDB; AAM49180, AAM49181.
XX
XX Novel DNA used as a capturing probe for detecting hepatitis B virus.
XX
XX Example 1; Page 14-16; 21pp; Japanese.
XX
XX The invention relates to nucleic acid sequences used to detect hepatitis
CC B virus (HBV). The invention specifically claims a sense capture probe
CC (ABL55199) corresponding to a fragment of the surface antigen (HBsAg)
CC coding sequence (bases 150-290 of the HBV subtype adr DNA shown in
CC ABL55204), and to inosine-containing primers (ABL55201-ABL55202) which
CC bind to regions of the HBV genome in or near the HBsAg coding sequence.
CC The invention also relates to DNA molecules in which at least one base is
CC deleted, replaced or added in the capture probe sequence, and which
CC hybridises to HBV DNA between bases 21-710 of ABL55204, and additionally
CC encompasses methods and reagents used in detecting HBV using the capture
CC probe of the invention. The present sequence represents a fragment of the
CC HBV subtype adr genome encoding the surface and core antigens used in an
CC exemplification of the invention
XX
SQ Sequence 3215 BP; 717 A; 860 C; 719 G; 919 T; 0 U; 0 Other;

Query Match 22.1%; Score 33.4; DB 6; Length 3215;
Best Local Similarity 55.7%; Pred. No. 0.46;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 18 TAAACAGGCCAAATGTTGCCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCCG 77
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1679 TGAACAGACCAATTATGCTTACAGCCTCCTAGTACAAAGACCATTAACTTCTCTC 1620

QY 78 ATCAACCTTCTCTGCCTATTAAATAAAATGTATCAAACTCATCCTGCAAGGTGG 132
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1619 CCCCAACTCCTCCAGTCTTTAAACAAACAGTCTTTGAAATATGCTTCAAGGTCTG 1565

RESULT 14
AAF22296
ID AAF22296 standard; DNA; 129021 BP.
XX
AC AAF22296;
XX
DT 20-MAR-2001 (first entry)
XX
DE BAC containing repeats from centromeres 1-4 #19.
XX
KW Centromere; michrosome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US007392.
XX
PR 18-MAR-1999; 99US-0125219P.
PR 01-APR-1999; 99US-0127409P.
PR 18-MAY-1999; 99US-0134770P.
```

```
PR 13-SEP-1999; 99US-0153584P.
PR 17-SEP-1999; 99US-0154603P.
PR 16-DEC-1999; 99US-0172493P.
XX
XX (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited michrosomes which can serve as vectors for the
PT construction of transgenic plant and animal cells.
XX
PS Claim 102; Page 686-716; 1449pp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for the
CC construction of transgenic plant and animal cells expressing selected
CC proteins such as hormones, enzymes, interleukins, clotting factors,
CC cytokines, antibodies, and growth factors
XX
SQ Sequence 129021 BP; 42091 A; 22610 C; 24583 G; 39735 T; 0 U; 2 Other;

Query Match 21.3%; Score 32.2; DB 3; Length 129021;
Best Local Similarity 59.1%; Pred. No. 5.4;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 27 CAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCCGATCAACCTT 86
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
124933 CTAGTGCTCTTTCAACTTTCAGTTCCAGTTGAAAGATCCGTGTACAGGAGACTGTAAACACTT 124992

QY 87 CTCTGCCTATTAAATAAAATGTCATAAACTCA 119
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
124993 CCATGCCGAGCAGAAATCAAATGTCATAAGCTGA 125025

RESULT 15
ABT23617/c
ID ABT23617 standard; DNA; 560 BP.
XX
AC ABT23617;
XX
DT 22-MAY-2003 (first entry)
XX
DE Stabilising reagent method related oligo SEQ ID No 69.
XX
KW Stabilising reaction reagent; PCR; primer; RNaseH; long-term storage;
KW specific amplification; pathogenic microorganism; chimeric;
KW genetic engineering; clinical medicine; ss.
XX
OS Hepatitis B virus.
XX
PN WO2002101042-A1.
XX
PD 19-DEC-2002.
XX
XX 12-JUN-2002; 2002WO-JP005832.
XX
PR 12-JUN-2001; 2001JP-00177737.
PR 20-AUG-2001; 2001JP-00249689.
XX
XX (TAKI ) TAKARA BIO INC.
XX
XX Sagawa H, Uemori T, Mukai H, Yamamoto J, Tomono J, Kobayashi E;
PI Enoki T, Asada K, Kato I;
XX
XX WPI; 2003-148805/14.
XX
XX Method for stabilizing and storing reaction reagents for specific
PT amplification and detection of nucleic acids particularly in e.g.
PT identifying pathogenic microorganisms or viruses in sample.
```

```
XX Example 15; Page 131-132; 177pp; Japanese.
PS
XX
CC The invention relates to a novel stabilising reaction reagent for use in
CC the amplification and/or detection of a target nucleic acid comprising:
CC preparing a reaction mixture with e.g. a nucleic acid as template, at
CC least 1 primer and RNaseH; and incubation of the reaction mixture for a
CC defined period of time to form a reaction product during the
CC amplification of such target nucleic acid. The method is useful for
CC stabilising and long-term storage of reaction reagents for highly
CC sensitive and specific amplification and detection of nucleic acids
CC particularly in identifying pathogenic microorganisms or viruses in a
CC sample using chimeric oligonucleotide primers, which is useful in genetic
CC engineering and clinical medicine. This polynucleotide sequence
CC represents an oligo relating to the novel stabilising reaction reagent
CC method of the invention
XX
SQ Sequence 560 BP; 101 A; 167 C; 143 G; 149 T; 0 U; 0 Other;

Query Match      21.1%; Score 31.8; DB 7; Length 560;
Best Local Similarity 54.8%; Pred. No. 0.84;
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 18 TAAACAGGCCAAATGTTGCCTTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCGC 77
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 TGAACAGACCAATTATGCCTACAGCCCTCCTAGTACAAAGACCTTTAAACCTAATCTCCTC 380

QY 78 ATCAACCTTCTCTGCCCTATTAAATAAATGTCATAAACTCATCCTGCAAGGTGG 132
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 CCCCAACTCCTCCAGTCTTTAAACAACACAGTCTTTGAAGTATGCCCTCAAGGTCG 325
```

Search completed: August 10, 2004, 21:54:08  
Job time : 64.5488 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 21:58:32 ; Search time 70.0822 Seconds  
(without alignments)  
10571.835 Million cell updates/sec

Title: US-10-001-857-41  
Perfect score: 151  
Sequence: 1 ccgccccggcaggtacctaa.....gcaaatctctcaagaatatg 151

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	151	14	US-10-001-857-41
2	140.4	93.0	3096	14	US-10-001-857-42
3	93.8	62.1	602	10	US-09-814-353-3218
4	93.8	62.1	602	10	US-09-814-353-9543
5	93.8	62.1	745	10	US-09-814-353-15927
6	84.8	56.2	2488	17	US-10-467-433-23
7	83.8	55.5	815	14	US-10-001-857-111
8	83.2	55.1	2610	10	US-09-814-353-20725
9	33.4	22.1	3215	13	US-10-453-792-296
10	33.4	22.1	3215	13	US-10-453-792-297
11	31.8	21.1	3212	13	US-10-453-792-298
12	31.8	21.1	3215	13	US-10-453-792-290
13	31.8	21.1	3215	13	US-10-453-792-291
14	31.8	21.1	3215	13	US-10-453-792-292

C 15	31.8	21.1	3215	13	US-10-453-792-295	Sequence 295, App
C 16	31.8	21.1	3215	13	US-10-449-801A-11	Sequence 11, Appl
C 17	31.4	20.8	3213	13	US-10-453-792-288	Sequence 288, App
C 18	31.4	20.8	3213	13	US-10-453-792-289	Sequence 289, App
C 19	31	20.5	1644	13	US-10-282-122A-23451	Sequence 23451, A
C 20	31	20.5	3161	13	US-10-453-792-301	Sequence 301, App
21	30.8	20.4	582	13	US-10-027-632-221960	Sequence 221960,
22	30.8	20.4	582	16	US-10-027-632-221960	Sequence 221960,
23	30.6	20.3	322101	13	US-10-354-247-1	Sequence 1, Appli
24	30.6	20.3	322101	15	US-10-060-902-1	Sequence 1, Appli
C 25	30.2	20.0	711	13	US-10-027-632-149134	Sequence 149134,
C 26	30.2	20.0	711	16	US-10-027-632-149134	Sequence 149134,
C 27	30.2	20.0	3214	13	US-10-453-792-294	Sequence 294, App
C 28	30.2	20.0	3215	13	US-10-453-792-286	Sequence 286, App
C 29	30.2	20.0	197526	17	US-10-322-281-498	Sequence 498, App
C 30	30	19.9	3480	13	US-10-027-632-114041	Sequence 114041,
C 31	30	19.9	3480	13	US-10-027-632-114042	Sequence 114042,
C 32	30	19.9	3480	16	US-10-027-632-114041	Sequence 114041,
C 33	30	19.9	3480	16	US-10-027-632-114042	Sequence 114042,
C 34	30	19.9	13273	9	US-09-764-869-2349	Sequence 2349, Ap
C 35	30	19.9	13273	15	US-10-091-504-2349	Sequence 2349, Ap
C 36	30	19.9	13273	16	US-10-227-577-2349	Sequence 2349, Ap
C 37	30	19.9	300000	15	US-10-262-552-33	Sequence 33, Appl
C 38	30	19.9	300000	17	US-10-703-210-33	Sequence 33, Appl
39	29.8	19.7	1510	13	US-10-424-599-125454	Sequence 125454,
40	29.6	19.6	576	13	US-10-282-122A-29360	Sequence 29360, A
C 41	29.6	19.6	1497	16	US-10-369-493-33434	Sequence 33434, A
C 42	29.6	19.6	4408	17	US-10-437-963-56946	Sequence 56946, A
C 43	29.4	19.5	197	9	US-09-837-297-2	Sequence 2, Appli
C 44	29.4	19.5	250	10	US-09-875-453-16	Sequence 16, Appl
45	29.4	19.5	258	13	US-10-424-599-63730	Sequence 63730, A

ALIGNMENTS

RESULT 1

US-10-001-857-41  
; Sequence 41, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41  
; LENGTH: 151  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-001-857-41

Query Match 100.0%; Score 151; DB 14; Length 151;  
Best Local Similarity 100.0%; Pred. No. 3.1e-42;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCTGTTTCAACAGC	60
Db	1	CCGCCCGGCGAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGTTCTGTTTCAACAGC	60
QY	61	ATGGTGTGAAGCGCCGCATCAACCTTCTGCCTATTATAAATAAAATGTCATAAACTCAT	120
Db	61	ATGGTGTGAAGCGCCGCATCAACCTTCTGCCTATTATAAATAAAATGTCATAAACTCAT	120
QY	121	CCTGCAAGGTGGCAAATTCCTCAAGATATG	151



```
Db      121 CCTGCAAGGTGGCAAATTCCTCAAGAATATG 151
|||||
RESULT 2
US-10-001-857-42/c
; Sequence 42, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 3096
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-42
Query Match      93.0%; Score 140.4; DB 14; Length 3096;
Best Local Similarity 99.3%; Pred. No. 6.4e-38;
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      10 CAGGTACCTAAACAGGCCAAATGTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 69
|||||
Db      2052 CAGGTACCTAAACAGGCCAAATGTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGGA 1993

QY      70 AGCGCGCATCAACCTTCTCGCTATTAAAAATAAAATGTCATAAACTCATCTGCAAGG 129
|||||
Db      1992 AGCGTGCATCAACCTTCTCGCTATTAAAAATAAAATGTCATAAACTCATCTGCAAGG 1933

QY      130 TGGCAAATTCCTCAAGAATATG 151
|||||
Db      1932 TGGCAAATTCCTCAAGAATATG 1911

RESULT 3
US-09-814-353-3218
; Sequence 3218, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 561_590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9543
Query Match      62.1%; Score 93.8; DB 10; Length 602;
Best Local Similarity 89.4%; Pred. No. 4.8e-22;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      1 CCGCCCGGCGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 60
|||||
Db      14 CCGCCCGGCGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 73

QY      61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCATA 113
|||||
Db      74 ATGGTGTGAAGCGCTGCATCAACCTTCTCTGCCTCATCTGCAAGGTGGCAA 126

RESULT 4
US-09-814-353-9543
; Sequence 9543, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 561_590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9543
Query Match      62.1%; Score 93.8; DB 10; Length 602;
Best Local Similarity 89.4%; Pred. No. 4.8e-22;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      1 CCGCCCGGCGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 60
|||||
Db      14 CCGCCCGGCGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGC 73

QY      61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAAAATAAAATGTCATA 113
|||||
Db      74 ATGGTGTGAAGCGCTGCATCAACCTTCTCTGCCTCATCTGCAAGGTGGCAA 126
```

RESULT 5  
US-09-814-353-15927  
; Sequence 15927, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 15927  
; LENGTH: 745  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 571  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-15927

Query Match 62.1%; Score 93.8; DB 10; Length 745;  
Best Local Similarity 89.4%; Pred. No. 5.3e-22;  
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
  
QY 1 CCGCCCGGGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCTGTGTTTCAACAGC 60  
Db 101 CCGCCCGGGCAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCTGTGTTTCAACAGC 160  
  
QY 61 ATGGTGTGAAGCGCGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
Db 161 ATGGTGTGAAGCGCTGCATCAACCTTCTCTGCCTCATCCTGCAAGTGGCAAA 213

RESULT 6  
US-10-467-433-23/c  
; Sequence 23, Application US/10467433  
; Publication No. US20040087773A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION;  
; APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;  
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;  
; APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;  
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;  
; APPLICANT: DING, Li; GIETZEN, Kimberly J.;  
; APPLICANT: ISON, Craig H.; LU, Dyung Aina M.;  
; APPLICANT: HAFALIA, April J.A.; GANDHI, Ameena R.;  
; APPLICANT: THANGAVELU, Kavitha; SANJANWALA, Madhusudan M.;  
; APPLICANT: TANG, Y. Tom; RAMKUMAR, Jayalaxmi;  
; APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;  
; APPLICANT: AZIMZAI, Yalda; SAPPERSTEIN, Stephanie K.;  
; APPLICANT: BURFORD, Neil; LEE, Ernestine A.;  
; APPLICANT: LU, Yan; TRAN, Uyen K.;  
; APPLICANT: MARQUIS, Joseph P.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT

; FILE REFERENCE: PF-0899 USN  
; CURRENT APPLICATION NUMBER: US/10/467,433  
; CURRENT FILING DATE: 2003-08-06  
; PRIOR APPLICATION NUMBER: PCT/US02/03709  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/268,117  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/269,618  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: US 60/271,118  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: US 60/274,486  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/274,436  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: US 60/334,229  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: US 60/353,284  
; PRIOR FILING DATE: 2002-02-01  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PERL Program  
; SEQ ID NO 23  
; LENGTH: 2488  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 7111920CB1  
US-10-467-433-23

Query Match 56.2%; Score 84.8; DB 17; Length 2488;  
Best Local Similarity 88.5%; Pred. No. 1.3e-18;  
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
  
QY 10 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCTGTGTTTCAACAGCATGGTGGA 69  
Db 1554 CAGGTACCTAAACAGGCCAAATGTTGCCTTTGGGGTTCTGTGTTTCAACAGCATGGTGGA 1495  
  
QY 70 AGCGCGGCATCAACCTTCTCTGCCTATTAAATAAAATGTCATA 113  
Db 1494 AGCGCTGCATCAACCTTCTCTGCCTCATCCTGCAAGTGGCAAA 1451

RESULT 7  
US-10-001-857-111  
; Sequence 111, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 111  
; LENGTH: 815  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (206)..(305)  
; OTHER INFORMATION: a, c, g or t  
US-10-001-857-111  
  
Query Match 55.5%; Score 83.8; DB 14; Length 815;  
Best Local Similarity 88.3%; Pred. No. 1.7e-18;

Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 11 AGGTACTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGAA 70  
Db 87 AGGTACTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGAA 146  
QY 71 GCGCCGCATCAACCTTCTCTGCCCTATTAAAAATAAAATGTGATA 113  
Db 147 GCGCTGCATCAACCTTCTCTGCCCTCATCTGCAAGGTGGCAA 189

RESULT 8  
US-09-814-353-20725  
; Sequence 20725, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20725  
; LENGTH: 2610  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1, 2, 3, 4  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20725

Query Match 55.1%; Score 83.2; DB 10; Length 2610;  
Best Local Similarity 87.5%; Pred. No. 4.7e-18;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 10 CAGGTACTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 69  
Db 1009 CAGGTACTAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGA 1068  
QY 70 AGCGCGCATCAACCTTCTCTGCCCTATTAAAAATAAAATGTGATA 113  
Db 1069 AGCGCTGCATCAACCTTCTCTGCCCTCATCTGCAAGGTGGCAA 1112

RESULT 9  
US-10-453-792-296/c  
; Sequence 296, Application US/10453792  
; Publication No. US20040029110A1  
; GENERAL INFORMATION:  
; APPLICANT: STUYVER, LIEVEN  
; APPLICANT: ROSSAU, RUDI  
; APPLICANT: MAERTENS, GEERT  
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV  
; NUMBER OF SEQUENCES: 313  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/453,792  
FILING DATE: 04-Jun-2003  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/155,885A  
FILING DATE: 08-Oct-1998  
APPLICATION NUMBER: PCT/EP97/02002  
FILING DATE: 21-APR-1997  
APPLICATION NUMBER: EP 96870053.4  
FILING DATE: 19-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 2551-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 296:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3215 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 296:  
US-10-453-792-296  
Query Match 22.1%; Score 33.4; DB 13; Length 3215;  
Best Local Similarity 55.7%; Pred. No. 1.1;  
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 18 TAAACAGGCCAAATGTTGCCTTTGGGTTCTCTGTTTCAACAGCATGGTGTGAAGCCGCG 77  
Db 1807 TGAACAGACCAATTTATGCTTACAGCTCCTAGTACAAAGACCATTAACCTAATCTCTC 1748  
QY 78 ATCAACCTTCTCTGCCTATTAAAAATAAAATGTGATAAACTCATCTCTGCAAGGTGG 132  
Db 1747 CCCCACTCTCCAGTCTTTAAACAAACAGTCTTTGAAATATGCTCAAGGTGCG 1693  
RESULT 10  
US-10-453-792-297/c  
; Sequence 297, Application US/10453792  
; Publication No. US20040029110A1  
; GENERAL INFORMATION:  
; APPLICANT: STUYVER, LIEVEN  
; APPLICANT: ROSSAU, RUDI  
; APPLICANT: MAERTENS, GEERT  
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV  
; NUMBER OF SEQUENCES: 313  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/453,792  
; FILING DATE: 04-Jun-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/155,885A  
; FILING DATE: 08-Oct-1998  
; APPLICATION NUMBER: PCT/EP97/02002  
; FILING DATE: 21-APR-1997  
; APPLICATION NUMBER: EP 96870053.4  
; FILING DATE: 19-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B.J.  
; REGISTRATION NUMBER: 36,663  
; REFERENCE/DOCKET NUMBER: 2551-5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 297:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3215 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; SEQUENCE DESCRIPTION: SEQ ID NO: 297:  
US-10-453-792-297

Query Match 22.1%; Score 33.4; DB 13; Length 3215;  
Best Local Similarity 55.7%; Pred. No. 1.1;  
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
  
QY 18 TAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCCGC 77  
Db 1807 TGAACAGACCAATTATGCTTACAGCCTCTAGTACAAAGACCAATTAACTTAATCTCTC 1748  
  
QY 78 ATCAACCTTCTCTGCCTATTAAATAAATGTATCAAACTCATCTGCAAGGTGG 132  
Db 1747 CCCCACTCTCCAGTCTTTAAACAAACAGTCTTTGAAATATGCCTCAAGGTGCG 1693

RESULT 11  
US-10-453-792-298/c  
; Sequence 298, Application US/10453792  
; Publication No. US20040029110A1  
; GENERAL INFORMATION:  
; APPLICANT: STUYVER, LIEVEN  
; ROSSAU, RUDI  
; MAERTENS, GEERT  
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV  
; NUMBER OF SEQUENCES: 313  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/453,792  
; FILING DATE: 04-Jun-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/155,885A  
; FILING DATE: 08-Oct-1998  
; APPLICATION NUMBER: PCT/EP97/02002  
; FILING DATE: 21-APR-1997  
; APPLICATION NUMBER: EP 96870053.4  
; FILING DATE: 19-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B.J.  
; REGISTRATION NUMBER: 36,663  
; REFERENCE/DOCKET NUMBER: 2551-5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 298:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3212 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; SEQUENCE DESCRIPTION: SEQ ID NO: 298:  
US-10-453-792-298

Query Match 21.1%; Score 31.8; DB 13; Length 3212;  
Best Local Similarity 54.8%; Pred. No. 4;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
  
QY 18 TAAACAGGCCAAATGTTGCCTTTGGGGTTCCTGTTTCAACAGCATGGTGTGAAGCGCCGC 77  
Db 1804 TGAACAGACCAATTATGCTTACAGCCTCTCTAGTACAAAGATCATTAACCTTAAGTTCTC 1745  
  
QY 78 ATCAACCTTCTCTGCCTATTAAATAAATGTATCAAACTCATCTGCAAGGTGG 132  
Db 1744 CCCCACTCTCCAGTCTCTTAACAAACAGTCTTTGAAATATGCCTCAAGGTGCG 1690

RESULT 12  
US-10-453-792-290/c  
; Sequence 290, Application US/10453792  
; Publication No. US20040029110A1  
; GENERAL INFORMATION:  
; APPLICANT: STUYVER, LIEVEN  
; ROSSAU, RUDI  
; MAERTENS, GEERT  
; TITLE OF INVENTION: METHOD FOR TYPING AND DETECTING HBV  
; NUMBER OF SEQUENCES: 313  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/453,792  
; FILING DATE: 04-Jun-2003  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/155,885A  
; FILING DATE: 08-Oct-1998  
; APPLICATION NUMBER: PCT/EP97/02002  
; FILING DATE: 21-APR-1997  
; APPLICATION NUMBER: EP 96870053.4  
; FILING DATE: 19-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B.J.









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 12336.3 Seconds  
(without alignments)  
10877.634 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttctcagaaactcccagg.....atttttgacaaaaaaaaa 3096

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3096	100.0	3096	6	AX523454	AX523454 Sequence
2	2417.2	78.1	2488	6	AX537452	AX537452 Sequence
3	1838.4	59.4	2552	10	AY102701	AY102701 Mus muscu
4	1828.2	59.1	2594	10	BC056435	BC056435 Mus muscu
5	1799.2	58.1	2543	10	AF272892	AF272892 Rattus no
6	1766.6	57.1	1839	9	AK025266	AK025266 Homo sapi
7	1264.8	40.9	2467	5	BC053286	BC053286 Danio rer
8	988	31.9	997	9	AK026296	AK026296 Homo sapi
9	885.4	28.6	1796	9	AK056059	AK056059 Homo sapi
10	717.4	23.2	719	6	AX664135	AX664135 Sequence
11	664.2	21.5	147876	2	AC102010	AC102010 Mus muscu
12	551.4	17.8	181864	9	AL353743	AL353743 Human DNA
13	440	14.2	440	6	AX778899	AX778899 Sequence
14	398.4	12.9	160115	9	AL161447	AL161447 Human DNA
15	321.4	10.4	446	9	HS3334821	AJ3334821 Homo sapi
16	299	9.7	743	9	HS3334413	AJ3334413 Homo sapi
17	236.4	7.6	642	9	HS342328	AJ342328 Homo sapi
18	236.4	7.6	742	9	HS3331811	AJ3331811 Homo sapi
19	233.2	7.5	640	9	HS342329	AJ342329 Homo sapi
20	233.2	7.5	652	9	HS342327	AJ342327 Homo sapi
21	217	7.0	2537	3	AY060948	AY060948 Drosophil
22	207.6	6.7	2565	3	AY089452	AY089452 Drosophil
23	207	6.7	320902	2	AL161456	AL161456 Homo sapi
24	182	5.9	489	11	G63197	G63197 SHGC-140830
25	177.8	5.7	34935	9	AL161453	AL161453 Human DNA
26	177.6	5.7	111554	2	AC143812	AC143812 Macaca mu
27	155.8	5.0	113836	10	AL669837	AL669837 Mouse DNA
28	153.2	4.9	253611	2	AC111473	AC111473 Rattus no
29	153.2	4.9	271861	2	AC108541	AC108541 Rattus no
30	153.2	4.9	281804	2	AC134869	AC134869 Mus muscu
31	147.2	4.8	201890	2	AC135469	AC135469 Mus muscu
32	140.4	4.5	151	6	AX523453	AX523453 Sequence
33	134.4	4.3	204131	2	AC128898	AC128898 Rattus no
34	130.4	4.2	3497	10	AY102702	AY102702 Mus muscu
35	90.6	2.9	16016	2	AC020508	AC020508 Drosophil
36	90.6	2.9	60135	2	AC091200	AC091200 Drosophil
37	90.6	2.9	164443	3	AC012167	AC012167 Drosophil
38	90.6	2.9	298827	3	AE003463	AE003463 Drosophil
39	87.8	2.8	815	6	AX523523	AX523523 Sequence
40	69.4	2.2	7218	6	I66494	I66494 Sequence 14
41	66.6	2.2	2000	6	AX655393	AX655393 Sequence
42	66.6	2.2	110737	2	AC011105	AC011105 Homo sapi
43	65.6	2.1	68328	2	AC137705	AC137705 Homo sapi
44	64.6	2.1	832	11	BV026280	BV026280 S212P6647
45	64.4	2.1	69373	2	AC101456	AC101456 Mus muscu

ALIGNMENTS

RESULT 1  
AX523454  
LOCUS AX523454  
DEFINITION Sequence 42 from Patent WO02064788.  
ACCESSION AX523454  
VERSION AX523454.1 GI:24412350  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.  
TITLE Compositions and methods relating to lung specific genes and proteins

AX523454  
Sequence 42 from Patent WO02064788.  
AX523454  
AX523454.1 GI:24412350  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.  
Compositions and methods relating to lung specific genes and proteins



JOURNAL	Patent: WO 02064788-A 42 22-AUG-2002;
Diadexus, Inc. (US)	
FEATURES	Location/Qualifiers
source	1. .3096
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
ORIGIN	
	Query Match 100.0%; Score 3096; DB 6; Length 3096;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 3096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 TTCTCAGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 60
Db	1 TTCTCAGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 60
Qy	61 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTGCTGCTCGCAACCACTAAGGTCACG 120
Db	61 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTGCTGCTCGCAACCACTAAGGTCACG 120
Qy	121 CAAACCTCCACGGTTTCCTTCCGCCCTTCGGCTCACCTTTCTAAGAAATTTCCAGAGGGCA 180
Db	121 CAAACCTCCACGGTTTCCTTCCGCCCTTCGGCTCACCTTTCTAAGAAATTTCCAGAGGGCA 180
Qy	181 GCGCAGACGGGGCGGGCTCTGAGACTCCGGGCTCCGGCTCCTTTCCGGGAACCGCCCACTA 240
Db	181 GCGCAGACGGGGCGGGCTCTGAGACTCCGGGCTCCGGCTCCTTTCCGGGAACCGCCCACTA 240
Qy	241 CCCAGACTCCGACAGAGGGTGAAGAAAGATAACTTCCGGTCTCGCGATCGTCTCTAATC 300
Db	241 CCCAGACTCCGACAGAGGGTGAAGAAAGATAACTTCCGGTCTCGCGATCGTCTCTAATC 300
Qy	301 TCGCGAGAAGAGAGCGCGCCGCATCGCCGAAACGGAGGCGGTGGCGAGGGGGGTG 360
Db	301 TCGCGAGAAGAGAGCGCGCCGCATCGCCGAAACGGAGGCGGTGGCGAGGGGGGTG 360
Qy	361 TGGCCGGGAGCGCGAAGTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGTCCCGG 420
Db	361 TGGCCGGGAGCGCGAAGTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGTCCCGG 420
Qy	421 GCATACGCATGCGTGACGGTCCCGTCCGGCTGGGCTGAGAGGGAGGGGCGGCGGCG 480
Db	421 GCATACGCATGCGTGACGGTCCCGTCCGGCTGGGCTGAGAGGGAGGGGCGGCGGCG 480
Qy	481 GCCAGGCGGCGTGGTTATTTCCGTGGTCCGACAGTGCCTGAGAGGGGCGGCGGCGGAC 540
Db	481 GCCAGGCGGCGTGGTTATTTCCGTGGTCCGACAGTGCCTGAGAGGGGCGGCGGCGGAC 540
Qy	541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAG 600
Db	541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAG 600
Qy	601 CTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGSGTGGACATTACCCCAAGAT 660
Db	601 CTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGSGTGGACATTACCCCAAGAT 660
Qy	661 TTTGAAGAAGCTTGTGCGAGATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGT 720
Db	661 TTTGAAGAAGCTTGTGCGAGATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGT 720
Qy	721 CTTTTTGAAGCCATGCTGCTATTTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780
Db	721 CTTTTTGAAGCCATGCTGCTATTTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780
Qy	781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840
Db	781 GGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840
Qy	841 AAAATTAAAGATCTCACCTTGCCTGAACTGATAGGATTAATGGATACATGTTTTTGGCTGT 900
Db	841 AAAATTAAAGATCTCACCTTGCCTGAACTGATAGGATTAATGGATACATGTTTTTGGCTGT 900

Qy	901 TTGATAACGTGGTTAGAAAGGCCATTCACTGGCACAGACAGATATTACGTGCCTTTACATT 960
Db	901 TTGATAACGTGGTTAGAAAGGCCATTCACTGGCACAGACAGATATTACGTGCCTTTACATT 960
Qy	961 CATAATCCAGACTTTTATAGAAATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020
Db	961 CATAATCCAGACTTTTATAGAAATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020
Qy	1021 ATCTGTGACATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTTT 1080
Db	1021 ATCTGTGACATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTTT 1080
Qy	1081 CAGTCAATGACTTATGGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC 1140
Db	1081 CAGTCAATGACTTATGGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC 1140
Qy	1141 ATGCTAAAAAGATGTGGAGGATGCATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAA 1200
Db	1141 ATGCTAAAAAGATGTGGAGGATGCATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAA 1200
Qy	1201 GGAGAAGAAAGAGATCCAGAACTTGAACCTAGAACACCAACAATGTTTAGCAGTATTTCAGC 1260
Db	1201 GGAGAAGAAAGAGATCCAGAACTTGAACCTAGAACACCAACAATGTTTAGCAGTATTTCAGC 1260
Qy	1261 AGAGTGAATTTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAAAGAGACC 1320
Db	1261 AGAGTGAATTTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAAAGAGACC 1320
Qy	1321 AGTCTGTTGCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT 1380
Db	1321 AGTCTGTTGCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT 1380
Qy	1381 CATAATTCAATGTCATATGTCATCCAGGCCAGAAATGATACATAAAAAGGAGATCATCCA 1440
Db	1381 CATAATTCAATGTCATATGTCATCCAGGCCAGAAATGATACATAAAAAGGAGATCATCCA 1440
Qy	1441 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500
Db	1441 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500
Qy	1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGCAAGATTAATAGATAGA 1560
Db	1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGCAAGATTAATAGATAGA 1560
Qy	1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTTACATTTGATCCTGGATTTTTTC 1620
Db	1561 ATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTTACATTTGATCCTGGATTTTTTC 1620
Qy	1621 TGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTC 1680
Db	1621 TGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTC 1680
Qy	1681 CTGGTGGATAACAAAAAGGTCCTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCA 1740
Db	1681 CTGGTGGATAACAAAAAGGTCCTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCA 1740
Qy	1741 CTTGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCAAGTGTCTTACCTATATAATAATC 1800
Db	1741 CTTGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCAAGTGTCTTACCTATATAATAATC 1800
Qy	1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTACTCCTGTGTTTCGGCCATTTCTGTAGTC 1860
Db	1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTACTCCTGTGTTTCGGCCATTTCTGTAGTC 1860
Qy	1861 TTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920
Db	1861 TTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920
Qy	1921 AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTAATAGGCAGAGAGGT 1980
Db	1921 AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTAATAGGCAGAGAGGT 1980
Qy	1981 TGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAAACCCCAAGGCAACATTTTGGCCTG 2040

Db 1981 ||||| TGATGCGCGCTTACACCATGCTGTGAAACAGGAACCCCAAGGCAACATTTGSCCTG 2040

QY 2041 TTTAGGTACCTGGGTCTTTACCATAAACCTTCGCATTATGATACAGTACCTTCTAAAGTGG 2100

Db 2041 TTTAGGTACCTGGGTCTTTACCATAAACCTTCGCATTATGATACAGTACCTTCTAAAGTGG 2100

QY 2101 CTTTGAATTGGAACCTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160

Db 2101 CTTTGAATTGGAACCTCTACAGTATGACAGTACTATTACATATATTGGTATCTCTCTGA 2160

QY 2161 ATTCCCTTACCGCATGGTTGATGTCACACATTGAGTCGTGCCGATGGCTCTCAAAATGGCAGA 2220

Db 2161 ATTCCCTTACCGCATGGTTGATGTCACACATTGAGTCGTGCCGATGGCTCTCAAAATGGCAGA 2220

QY 2221 GGAAGGATAAATGGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACAAGAAAAAAA 2280

Db 2221 GGAAGGATAAATGGAAGAGCAGCAGAAAGCCGCTAGTAGTAAAAAACAAGAAAAAAA 2280

QY 2281 GAAAGTTCCGCCCATTTGAGCCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGC 2340

Db 2281 GAAAGTTCCGCCCATTTGAGCCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGC 2340

QY 2341 TGGAAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAACCGAAGTT 2400

Db 2341 TGGAAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAACCGAAGTT 2400

QY 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGAT 2460

Db 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGAT 2460

QY 2461 GACCCCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCGACCTCAATAAATATAG 2520

Db 2461 GACCCCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCGACCTCAATAAATATAG 2520

QY 2521 CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAAGCACTTCAACAGGCATA 2580

Db 2521 CCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAAGCACTTCAACAGGCATA 2580

QY 2581 AATGATATGGAATAATTCCTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640

Db 2581 AATGATATGGAATAATTCCTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640

QY 2641 CAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT 2700

Db 2641 CAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT 2700

QY 2701 TCCTCCTGAATTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTTTGAGA 2760

Db 2701 TCCTCCTGAATTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTTTGAGA 2760

QY 2761 GAGACTGGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG 2820

Db 2761 GAGACTGGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG 2820

QY 2821 CACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACTCATTA 2880

Db 2821 CACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACTCATTA 2880

QY 2881 TAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940

Db 2881 TAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940

QY 2941 TGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATTGATGAACGTTATAT 3000

Db 2941 TGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATTGATGAACGTTATAT 3000

QY 3001 GGTTTATTACAGATTTAATCACAAATCATTTTTTATGAATGATGAGTGAAAAATAGTGT 3060

Db 3001 GGTTTATTACAGATTTAATCACAAATCATTTTTTATGAATGATGAGTGAAAAATAGTGT 3060

QY 3061 TTATAAAGGTTAATAAATTTCTTGACAAAAA 3096

Db 3061 TTATAAAGGTTAATAAATTTCTTGACAAAAA 3096

RESULT 2

AX537452

LOCUS AX537452 2488 bp DNA linear PAT 23-NOV-2002

DEFINITION Sequence 23 from Patent WO02070709.

ACCESSION AX537452

VERSION AX537452.1 GI:25269238

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Lal,P.G., Baughn,M.R., Yao,M.G., Walia,N.K., Elliot,V.S., Xu,Y., Honchell,C.D., Yue,H., Ding,L., Gietzen,K.J., Ison,C.H., Lu,D.A., Hafalia,A.J., Ghandi,A.R., Thangavelu,K., Sanjanwala,M.M., Tang,Y.T., Ramkumar,J., Griffiin,J.A., Swarnaker,A., Azimzai,Y., Sapperstein,S.K., Burford,N., Lee,E.A., Lu,Y., Tran,U.K. and Marquis,J.P.

TITLE Molecules for disease detection and treatment

JOURNAL Patent: WO 02070709-A 23 12-SEP-2002;

INCYTE Genomics, Inc. (US)

FEATURES

Location/Qualifiers

1..2488

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

/note="Incyte ID No: 7111920CB1"

ORIGIN

Query Match 78.1%; Score 2417.2; DB 6; Length 2488;

Best Local Similarity 98.5%; Pred. No. 0;

Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;

QY 474 GGCGGCGCGAGGGCGTCTGTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGGGT 533

Db 1 GGCGGCGCGAGGGCGTCTGTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGGGT 60

QY 534 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGG 593

Db 61 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGG 120

QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTAC 653

Db 121 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAGCAATACAAACTGGGTGGACATTAC 180

QY 654 CCAAGATTTTGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCT 713

Db 181 CCAAGATTTTGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCT 240

QY 714 ATTTGGTCTTTTGAAGCCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGG 773

Db 241 ATTTGGTCTTTTGAAGCCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGG 300

QY 774 CATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGG 833

Db 301 CATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGG 360

QY 834 CACTATTAAAAATTAAAGATCTCACCTTGCCCTGAACTGATGAGGGATTATGGATACATGTTT 893

Db 361 CACTATTAAAAATTAAAGATCTCACCTTGCCCTGAACTGATGAGGGATTATGGATACATGTTT 420

QY 894 TTGCTGTTTGATAACGTGGTTAGAAAGGCCATTTCACCTGGCACAGACAGTATTACGTGCCT 953

Db 421 TTGCTGTTTGATAACGTGGTTAGAAAGGCCATTTCACCTGGCACAGACAGTATTACGTGCCT 480

QY 954 TTACATTTCATAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAAT 1013

Db 481 TTACATTTCATAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAAT 540

QY 1014 CTTGAAAAATCTGTGACATTGCAAGGGAAAAAGTAATAAAGCTGCTGTTTTTGAAGAGGA 1073



Db 541. CTTGAAAACTGTGACATTGCAAGGGAAGAAAGTAAATAAAGTGCTGTTTTTGAAAGAGGA 600  
QY 1074 AGATTTCAGTCAATGACTTATGGATTAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133  
Db 601 AGATTTCAGTCAATGACTTATGGATTAAATGGCTAACAGTGTGACAGATCTTCGAGT 660  
QY 1134 TACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAAGAGTACTCGAAG 1193  
Db 661 TACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAAGAGTACTCGAAG 720  
QY 1194 TCGACAAAGGAGAAAGAGATCCAGAAGTTGAAC TAGAACACCAACAATGTTTAGCAGT 1253  
Db 721 TCGACAAAGGAGAAAGAGATCCAGAAGTTGAAC TAGAACACCAACAATGTTTAGCAGT 780  
QY 1254 ATTACAGCAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCCTTATAGCCTTTACTAAGAA 1313  
Db 781 ATTACAGCAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCCTTATAGCCTTTACTAAGAA 840  
QY 1314 AGACACCAGTGTGTTGCAGAAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTC 1373  
Db 841 AGACACCAGTGTGTTGCAGAAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTC 900  
QY 1374 TGCCATTCAATAATTCAATTGCATCATGGCATCCAGGCCAGAGATGATACATAAAAAGGAGA 1433  
Db 901 TGCCATTCAATAATTCAATTGCATCATGGCATCCAGGCCAGAGATGATACATAAAAAGGAGA 960  
QY 1434 TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTCCACCTACCTT 1493  
Db 961 TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTCCACCTACCTT 1020  
QY 1494 CCTCGATATGCAAAAAATAATTAAAAGGGAAGAAATGGTGAATTAATTTTGAAGATTAAAT 1553  
Db 1021 CCTCGATATGCAAAAAATAATTAAAAGGGAAGAAATGGTGAATTAATTTTGAAGATTAAAT 1080  
QY 1554 AGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCTGGA 1613  
Db 1081 AGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCTGGA 1140  
QY 1614 TTTTCTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1673  
Db 1141 TTTTCTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1200  
QY 1674 CACTTTCCTGGTGATACAAAAAGGTCTTTTGGAAC TCACTCATGCAAGACATGGTGAA 1733  
Db 1201 CACTTTCCTGGTGATACAAAAAGGTCTTTTGGAAC TCACTCATGCAAGACATGGTGAA 1260  
QY 1734 AGATGCATTCGGTCTTTTGTGAGATCCTCCGAGTGCCTTCCGCCAAGTGCTACCTATAT 1793  
Db 1261 AGATGCATTCGGTCTTTTGTGAGTCTCTCG - -GTGCTTTCCGCCAAGTGCTACCTATAT 1318  
QY 1794 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTC 1853  
Db 1319 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTC 1378  
QY 1854 TGTAGTCTTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAAT 1913  
Db 1379 TGTAGTCTTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAAT 1438  
QY 1914 ATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAG 1973  
Db 1439 ATTCTTGAGGAATTTGCCACCTTGCAGGATG- - - - -AGGCAG 1475  
QY 1974 AGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 2033  
Db 1476 AGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 1535  
QY 2034 TGGCCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGATATGATACAGTACCTTC 2093  
Db 1536 TGGCCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGATATGATACAGTACCTTC 1595  
QY 2094 TAAGTGGCTTTGAATTGGAACCTTACAGTATGCACGAGTACTATTACATATATTGGTATC 2153

Db 1596 TAACTGGCTTTGAATTGGAAC TCTACAGTATGCACGAGTACTATTACATATATTGGTATC 1655  
QY 2154 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTTGAGTCGTGCCGATGGCTCTCAAA 2213  
Db 1656 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTTGAGTCGTGCCGATGGCTCTCAAA 1715  
QY 2214 TGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAAACAAAGA 2273  
Db 1716 TGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAAACAAAGA 1775  
QY 2274 AAAAAAGAAAGTTTCGCCCATTTAGCCGAGAGATCAAAATGAGCCAAAGCATATCAGAAAC 2333  
Db 1776 AAAAAAGAAAGTTTCGCCCATTTAGCCGAGAGATCAAAATGAGCCAAAGCATATCAGAAAC 1835  
QY 2334 TGTGTGCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAC 2393  
Db 1836 TGTGTGCTGGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAC 1895  
QY 2394 CGAAGTTTGAGCTTGATAGTGAACAAAGTTCCGGTATGAACACAGAGTTTGCTCCATTCAACA 2453  
Db 1896 CGAAGTTTGAGCTTGATAGTGAACAAAGTTCCGGTATGAACACAGAGTTTGCTCCATTCAACA 1955  
QY 2454 GTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA 2513  
Db 1956 GTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA 2015  
QY 2514 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACITTTCAAC 2573  
Db 2016 AATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACITTTCAAC 2075  
QY 2574 AGGCAAAAATGATATTGGAATAATTTCTTAACCCGGACCATGAGGTTAATAGAATTTTAA 2633  
Db 2076 AGGCAAAAATGATATTGGAATAATTTCTTAACCCGGACCATGAGGTTAATAGAATTTTAA 2135  
QY 2634 AGGTTGCCAAACCCCACTTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2693  
Db 2136 AGGTTGCCAAACCCCACTTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAT 2195  
QY 2694 CTAAAGTTCTCCTGAATTTGAATTTCTCTGCTCATAAATFAATTTCTGTTGTGAAACTTG 2753  
Db 2196 CTAAAGTTCTCCTGAATTTGAATTTCTCTGCTCATAAATFAATTTCTGTTGTGAAACTTG 2255  
QY 2754 TTTGAGAGAGACTGGGGAGGTGCCATAAAGGGCAGAGTCTTCTTTCAGACCCCACTCT 2813  
Db 2256 TTTGAGAGAGACTGGGGAGGTGCCATAAAGGGCAGAGTCTTCTTTCAGACCCCACTCT 2315  
QY 2814 TAGAGGCACATCACCCAGGCTCCACATCACCGGAACTGAGATGGATTTCTTGGGTAACAA 2873  
Db 2316 TAGAGGCACATCACCCAGGCTCCACATCACCGGAACTGAGATGGATTTCTTGGGTAACAA 2375  
QY 2874 CTCATTATAAGGAATACCTTTAGTTTGACAGCCTTATATGACATGAATGAAAAC TGTGT 2933  
Db 2376 CTCATTATAAGGAATACCTTTAGTTTGACAGCCTTATATGACATGAATGAAAAC TGTGT 2435  
QY 2934 TTTAAAGTGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAAT 2981  
Db 2436 TTTAAAGTGGTTTATTATGTTCCATGTAAGACACTGGGTTCCATTAAT 2483

RESULT 3  
AY102701  
LOCUS AY102701 2552 bp mRNA linear ROD 23-JUN-2002  
DEFINITION Mus musculus embryonic growth-associated protein EGAP mRNA,  
complete cds.  
ACCESSION AY102701  
VERSION AY102701.1 GI:21539895  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2552)  
AUTHORS Wenzlau,J.M. and Weiser-Evans,M.C.M.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Mouse Embryonic Growth Associated Protein (EGAP)									
	Unpublished									
	2 (bases 1 to 2552)									
	Wenzlau,J.M. and Weiser-Evans,M.C.M.									
	Direct Submission									
	Submitted (13-MAY-2002) Pediatrics, University of Colorado HSC,									
	4200 East Ninth Ave, Denver, CO 80262, USA									
	Location/Qualifiers									
	1. .2552									
	FEATURES	source								
/organism="Mus musculus"										
/mol_type="mRNA"										
/db_xref="taxon:10090"										
/chromosome="11"										
/dev_stage="embryo"										
55. .2232										
/codon_start=1										
/product="embryonic growth-associated protein EGAP"										
/protein_id="AAM52342.1"										
/db_xref="GI:21539896"										
CDS	/translation="MVMKATVDDDASGWELGVPEKMEKSSTSVDITQDFEDACRELK LGEILLHDKLFLGLFEAMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLSL PELIGIMDTFCCLITWLEHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICDIA REKYNKAAVFEEDFQSMYTGFKMANSVTDLRVTGMLKDVEDDLQRRVKSTRSRQEE RDPEVELEHQQLAAAFSRVKFTRVLLTLLIAFTKETSAVAEAQKLMVQAADLLSAIH TSLHHGIIQAQNGTTKGDHPIMMGFEPLVNQRLLPPTFPRIYAKIIKREEMVNYFSRLID RIKIVCEVNVNLPNLHCILDFFCFESESPCVLSRSLLOTTFLVDNKKVFGTHLMQDMV KDALRSFVSPVLPSPKCCLYNNHQAKDCIDSFVTHCVRPFCSLVQIHGHNRRQRDKL GHILFEFATLQDEAEKVDAALHTMLLKQEPQRQHLACGLTWVLYHNLRIIMIQLYLLSGF ELELYSMHEYIYIYWLSEFLYAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKTKK KKVRPLSREITMSQAYQNMCAGMFKTMVAFDMGDKVRKPKFELDSEQVRYEHRFAPFN SVMTPPPVHYLQFKEMSDLNKYSPPPSQPELYVAASKHFQQTCKMILENIPNDREVS ILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSVHKYFPVVVKLV"									
	ORIGIN									
	Query Match 59.4%; Score 1838.4; DB 10; Length 2552;									
	Best Local Similarity 84.4%; Pred. No. 0;									
	Matches 2160; Conservative 0; Mismatches 341; Indels 57; Gaps 6;									
	QY	541	GGAGAAGTAGGCATAAATGGTTATGAAAGCTTCTGTAGATGATGACGATTGAGGATGGGAG	600						
	Db	40	GGCGCGACGGCATCATGTTATGAAAGCTACTGTAGACGACGATGCTTCGGGATGGGAG	99						
	QY	601	CTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAGAT	660						
	Db	100	CTCGGGTCCCGAAAAAATGGAAAAAGTAGCACAAAGCTGGGTGGACATAACCCAGGAT	159						
	QY	661	TTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTGTTGGT	720						
Db	160	TTTGAAGATGCTTGTGAGAGCTGAAGTTGGGAGAACTGCTTCATGATAAGCTGTTTGGT	219							
QY	721	CTTTTGAAGCCATGTCTGCTATTGAATGATGATCCCAAGATGGATGCTGGCATGATT	780							
Db	220	CTTTTGAAGCCATGTCTGCTATTGAATGATGATCCTAAGATGGATGCCGATATGATC	279							
QY	781	GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAAGCTATCAAGGATGGCATT	840							
Db	280	GGGAACCAAGTGAATAGAAAAGTTCTCAATTTTGAACAAAGCTATCAAGGATGGCACCATT	339							
QY	841	AAAATTAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTTTGTCTGT	900							
Db	340	AAAATTAAGACCTCAGCCTGCCTGAACTGATAGGAATAATGGACACCTGTTTCTGTCTGT	399							
QY	901	TTGATAACGTGGTTAGAAAGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATT	960							
Db	400	TTGATCATATGGCTCGAAGGCCATTCTTGGCACAGACAGTGTTTACGTGCCTTTACATT	459							
QY	961	CATAATCCAGACTTTATAGAAGATCCTGCTATGAAGCTTTTGTCTCTGGGAATCTTGAAA	1020							
Db	460	CATAATCCGACTTCATAGAAGATCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTGAAG	519							
QY	1021	ATCTGTGACATTCGAAGGGAAAAAGTAATAAAGCTGCTGTTTTTGAAGAGGAATTTT	1080							
Db	520	ATCTGCGACATTGCACGGGAAAAAGTAATAAAGCTGCTGTTTTTCGAAGAGGAAGATTTC	579							

QY	1081	CAGTCAATGACTTATGGATTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGC	1140			
Db	580	CAGTCAATGACATACGGATTAAAAATGGCAACACAGTGTGACAGATCTCCGAGTTACAGGG	639			
QY	1141	ATGCTAAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAGTCGACAA	1200			
Db	640	ATGCTAAAAGGATGTGGAAGATGATCTGCAAGGCGAGTAAAGAGTACTCGAAGTCGACAA	699			
QY	1201	GGAGAAGAAAGAGATCCAGAAAGTTGAAC TAGAACCAACCAACAATGTTTAGCAGTATTTCAGC	1260			
Db	700	GGAGAAGAGAGAGATCCAGAGGTCGAACTAGAACACCAAGCAGAGTGCCTGGCAGCATTCAGC	759			
QY	1261	AGAGTGAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACC	1320			
Db	760	AGAGTGAAGTTCACGCGAGTGTCTGACACTGCTCATAGCCTTTTACTAAGAAAAGAGACC	819			
QY	1321	AGTGTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATT	1380			
Db	820	AGTGTGTTGCAGAGGCTCAGAAACTGATGGTCCAGGCGGAGACCTTCTTCTGCCATT	879			
QY	1381	CATAATTCAATTGCATATGGCATCCAGGCCAGAAATGATCTACAAAAGGAGATCATCCA	1440			
Db	880	CACACCTCATTCACACCGGCATCCAGGCTCAGAATGGCACTACCAAAGGAGACCATCCA	939			
QY	1441	ATTATGATGGGTTTTGAACCCCTTGTGAACCAAGAGGCTACTTCCACCTACCTTCCCTCGA	1500			
Db	940	ATTATGATGGGTTTTGAGCCCCCTTGTTAAACCAAGAGATTACTTCCACCCACCTTCCCTCGC	999			
QY	1501	TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAAC TATTTTGCAGATTAATAGATAGA	1560			
Db	1000	TATGCAAAAAATAATTAAAGAGAGAAATGGTCAACTATTTCTCAAGATTAATAGACAGA	1059			
QY	1561	ATAAAACTGTCTGTGAGGTTGTGAATTTAAACAAATTTACATTTATTCCTGGATTTTTC	1620			
Db	1060	ATAAAACTGTCTGTGAAGTCGTGAAC TTAACAAACTTACACTGTATCCTGGATTTTTC	1119			
QY	1621	TGTGAATTTAGTGAACAGTCACCATGTGTCTTTCAAGATCTCTGTTACAAACCACTTTC	1680			
Db	1120	TGTGAATTTAGTGAGCAGTCACCATGTGTCTTTCAAGATCTCTGTTACAAACCACTTTC	1179			
QY	1681	CTGTGGATAACAAAAAGGCTCTTTGGAAC TCACTCATGCAAGACATGGTGAAAGATGCA	1740			
Db	1180	CTTGTGGATAACAAAAAGGCTCTTTGGCACTCATCTCATGCAAGACATGGTGAAAGATGCT	1239			
QY	1741	CTTCGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCCCAAGTGTACCTATATAATAATC	1800			
Db	1240	CTGCGGTCTTTCGTGAG - -TCCTCCGTGCTCTCCCCCAAGTGTGCTTATATAATAATC	1297			
QY	1801	ACCAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTTCGGCCATTCTGTAGTC	1860			
Db	1298	ACCAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTTCGGCCGTTTGTAGTC	1357			
QY	1861	TTATTGAGATCCATGGACATACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG	1920			
Db	1358	TTGTTGAGATCCATGGCCATACAGGGCTCGGACAGAGAGACAAGCTTGGTCACATTCTTG	1417			
QY	1921	AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGGT	1980			
Db	1418	AGGAGTTTGCTACCTTGCAAGATG-----AGGCAGAGAAGGT 1454				
QY	1981	TGATGACAGCGCTTCACACCATGTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTG	2040			
Db	1455	TGATGACAGCGCTTCATACTATGTTGTTGAAACAGGAGCCTCAGAGACAACATCTGGCCTG	1514			
QY	2041	TTTAGTACCTGGTCCCTTTACATAACCTTCGCATTATGATACAGTACCTTCTAAGTG	2100			
Db	1515	CTTAGGAACCTGGGTTCTTTACCATAACCTCCGGATTATGATCCAGTATCTGCTCAGTG	1574			
QY	2101	CTTTGAATTTGAACTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGA	2160			
Db	1575	CTTTGAGCTGGAGCTGTACAGCATGCATGAGTACTACTACATCTACTGTTACCTCTCCGA	1634			



QY	2161	ATTCTTTTACGCATGTTGATGTC AACATTGAGTCGTGCGGATGGCTCTCAAATGGCAGA	2220
Db	1635	GTTCCTGTATGTCATGGCTGATGTCAACCCCTGAGCCGCGCTCAGATGGCAGA	1694
QY	2221	GGAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAATAAAACAAAGAAAAAA	2280
Db	1695	GGAGAGGATCATGGAGGAACAGCAGAAAGGCCGCGCAGCAGCAAAAAACAAAGAGAAAA	1754
QY	2281	GAAAGTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCAGCAATATCAGAACATGTGTGC	2340
Db	1755	GAAAGTTCGTCCGTTGAGCCGAGAGATCACAATGAGCCAGCCGTATCAGAACATGTGTGC	1814
QY	2341	TGGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTT	2400
Db	1815	TGGGATGTTTCAAGACGATGGTCGCTTTTGACATGGATGGCAAGTGGCAAGCCCAAGTT	1874
QY	2401	TGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCANTCAACAGTGTGAT	2460
Db	1875	TGAGCTTGACAGTGAGCAGGTCCGGTATGAGCACAGGTTTGACCGTTCAACAGTGTGAT	1934
QY	2461	GACCCCGCGCCAGTGCACCTACTTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG	2520
Db	1935	GACGCGGCACCACTGCACCTACCTGCAGTTCAAGGAAATGTCTGACCTCAATAAATACAG	1994
QY	2521	CCCTCCTCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCACACAGGCAAA	2580
Db	1995	CCCTCCTCTCAGTCTCCAGAGCTGTATGTGGCAGCTAGTAAGCACTTTCAGCAGACAAA	2054
QY	2581	AATGATATTTGGAATAATATTCCTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGTTGC	2640
Db	2055	AATGATCCTTGAATAATATCCCAATCCAGACCGTGAGGTCAGTAGAATCCTGAAGTTGC	2114
QY	2641	CAAAACCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT	2700
Db	2115	CAAGCCCAACTTTGTGGTTATGAAGCTCTTAGCAGGAGGACACAAAAAGAGTCAAAGGT	2174
QY	2701	TCCTCCTGAAATTTGATTTCTCTGCTCAFAAATATTTTCCTGTTGTGAAACTTGTGTGAGA	2760
Db	2175	TCCTCCCGAATTCGACTTCTCTGTTCAAAAATATTTTCCTGTTGTGAAACTGTTTGAAA	2234
QY	2761	GAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCAACTCTTAGAGGG	2820
Db	2235	GAGACTGCAAGATGACCGTGTCCACTC-----TAGGGAA	2269
QY	2821	CACATCACCAGGCTCCACATCAGC--GGAAGTGAGATGGATTTCTTGGGTAACAACCTCAT	2878
Db	2270	GACATCAGTCTGCCACACATGGTGACAGAAAGTGGATCTCTTGAATGACAGACAGCTTGC	2329
QY	2879	TATAAGGAATACTTTTAGTTTGACAGCCCTATATGACATGAATGAAAACTGCTGTTTTAA	2938
Db	2330	TACAGGAAGACTTTTCAGTTTACCAACCTTACCTAGAGAGGATGAGAACTGCTGTTTTAA	2389
QY	2939	AGTGGTTTATATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATGTAACGTTAT	2998
Db	2390	AGTGGTTTATTGTACTCCATGGGTGTAAGTGG----CTGGAATGCATTTGGTGAACGCTAC	2445
QY	2999	ATGGTTTATTACAGATTTTAATCACAATCATTTTTTANGAATGATGAGTGAAAAATAGT	3058
Db	2446	ATGGTTTGATTACAGACTTAATTGTAAACATTTTTTTAATGAATGATGTGTG-APATAGT	2504
QY	3059	GTTTATAAAGGTTAATAAATTTCTTGACAAAAAATAAA 3096	
Db	2505	GTTTGTAAGGTTAATAAATTTCTTGACAAAAAATAAA 2542	
RESULT 4			
BC056435			
LOCUS			
DEFINITION	BC056435	2594 bp	linear
	Mus musculus RIKEN cdna C030004C14	gene, mRNA	CDNA clone MGC:56717
	IMAGE:5707801), complete cds.		
ACCESSION	BC056435		
VERSION	BC056435.1	GI:33989782	
KEYWORDS	MGC.		

SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE AUTHORS	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2594) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 22388257 12477932 2 (bases 1 to 2594) Strausberg,R. Direct Submission Submitted (14-AUG-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: M. Bento Soares, University of Iowa cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeeadi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
REMARK COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 124 Row: k Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21644570. Location/Qualifiers 1. .2594 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="MGC:66717 IMAGE:5707801" /tissue_type="Brain, mouse 15.5 dpc" /clone_lib="NIH_BMAP_EV0" /lab_host="DH10B" /note="Vector: pYX-ASC" 1. .2594 /gene="C030004C14Rik" /db_xref="LocusID:78689" /db_xref="MGI:1925939"
FEATURES	source
	gene





Db 1682 GGAGAGGATCATGGAGGAACAGCAGAAAGGCCGCGAGCCAGCAAAAACAAAGAGAAAA 1741

Qy 2281 GAAAGTTCGCCCATTTGAGCCGAGAGATCACAAATGAGCCAGCATATATCAGAACATGTGTGC 2340

Db 1742 GAAAGTTCGTCGGTTGAGCCGAGAGATCACAAATGAGCCAGGCTATCAGAACATGTGTGC 1801

Qy 2341 TGGATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAACCGAAGTT 2400

Db 1802 TGGGATGTTCAAGACGATGGTCGCTTTTGACATGGAATGGCAAAAGTCGCAAGCCCAAGTT 1861

Qy 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCANTCAACAGTGTGAT 2460

Db 1862 TGAGCTTGACAGTGAGCAGGTCCGGTATGAGCACAGGTTTGCACCGTTCAACAGTGTGAT 1921

Qy 2461 GACCCGCGCCAGTGCACCTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG 2520

Db 1922 GACGCGCCACAGTGCACTACCTGCAGTTCAGGAAATGTCTGACCTCAATAAATACAG 1981

Qy 2521 CCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTCAACAGGCAAA 2580

Db 1982 CCCTCCTCCTCAGTCTCCAGAGCTGTATGTGGCAGCTAGTAAGCACTTTCAGCAGGCAAA 2041

Qy 2581 AATGATATTGGAAAAATATTCTTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640

Db 2042 AATGATCCTTGAAAAATATCCCAATCCAGACCGTGAGGTCAGTAGAATCCTGAAGTTGC 2101

Qy 2641 CAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT 2700

Db 2102 CAAGCCCAACTTTGTGTTATGAAGTCTTTAGCAGGAGGACACAAAAAGAGTCAAAGGT 2161

Qy 2701 TCCTCCGAAATTTGATTTCTCTGCTCAFAAATATTTTCCCTGTTGTGAAACTTGTGAGA 2760

Db 2162 TCCTCCCGAATTCGACTTCTCTGTTACAAAATATTTTCCCTGTTGTGAAACTGTTTGAAA 2221

Qy 2761 GAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCAACTCTTAGAGGG 2820

Db 2222 GAGACTGCAAAAGATGACCGTGTCCACTC-----TAGGGAA 2256

Qy 2821 CACATCACAGGCTCCACATCAGC--GGAAGTGAGATGGATTTCTTGGGTAACAACTCAT 2878

Db 2257 GACATCAGTCTGCCACACATGGTGACAGAAAGTGGATCTCTTGAATGACAGACAGCTTGC 2316

Qy 2879 TATAAGGAATACTTTT-AGTTTGACAGCCCTTATATGACATGAATGAAAACTGCTGTTTTA 2937

Db 2317 TACAAGGAAGACTTTTTCAGTTTACCACCCCTTACCTAGAGAGGATGAGAACTGCTGTTTTA 2376

Qy 2938 AAGTGGTTTATATGTTCCATGGAAGAACTGGTCTTATTGAATGCAATGATGAACGTTA 2997

Db 2377 AAGTGGTTTATGTACTCCAATGGGTAACCTGGG---CTGGAATGCAATGGTGAACGCTA 2432

Qy 2998 TATGGTTTATATACAGATTTAATCACAATCATTTTTTATGAATGAATGAGTGAATAATAG 3057

Db 2433 CATGGTTTGATACAGACTTAATTGTAAACATTTTAAATGAATGATGAGTGAATAATAG 2492

Qy 3058 TGTTTATAAGGTTAAATAAATTTCTTGACAAAAA 3096

Db 2493 TGTTTGTAAAGGTTAATAAATTTCTTGACAAAAA 2531

RESULT 5

AF272892

LOCUS AF272892 2543 bp mRNA linear ROD 15-SEP-2000

DEFINITION Rattus norvegicus corneal wound healing related protein mRNA, complete cds.

ACCESSION AF272892

VERSION AF272892.1 GI:8926319

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 2543)

AUTHORS Yi,X.J., Li,X.F. and Yu,F.S.

TITLE A novel epithelial wound-related gene is abundantly expressed in developing rat cornea and skin

JOURNAL Curr. Eye Res. 20 (5), 430-440 (2000)

MEDLINE 20314365

PUBMED 10855038

REFERENCE 2 (bases 1 to 2543)

AUTHORS Li,X. and Yu,F.

TITLE Direct Submission

JOURNAL Submitted (26-MAY-2000) The Schepens Eye Research, 20 Staniford Street, Boston, MA 02114, USA

FEATURES

source Location/Qualifiers

1..2543

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/db\_xref="taxon:10116"

64..2241

/codon\_start=1

/product="corneal wound healing related protein"

/protein\_id="AAF81791.1"

/db\_xref="GI:8926320"

/translation="MVNKAAVDDDDASGWELNVPEKMEKSSTSWVDITQDFEDACRELKLGELLHDKLFGLFEMASIEIMDDPKMDAGMIGNQVNRKVLNFEQAVKDGTIKIKDLSLPELIGIMDTFCCLITWLEGHSLAQTFTCLYIHNPDFIEDPAMKAFALGILKICDIAREKVNKAAAFEEEDFQSMTYGFKMANGVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQCLAAFSRVKFTRVLLTVLIAFTKKETSAVAEAQKLMVQAADLLLSAIHTSLHHGIIQAONGTTKGDPHIMMGFEPLVNQRLLPPTFRYAKIIKREEMVNYFSRLIDRIKTVCVVNLPNLHCILDFFCEFSQPCVLSRLIQTTFLDNKKVFGVTHLMQDMVKDAPRSFVSPVLPKCLYNNHOAKDCIDSFVTHCVRFCSLVQIHGHNRRARQDKLGHILEEFATLQDEAEKVDAAALHTMLLKQEPORQHACLGTWVLYHSLRIMIQYLLSGFDLELYSMHEYIYYIWSLEFLYAWLMSTLSRADGSGQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCAGMFKTMVAFDMGDKVRKPKFELDSEQVRYEHRFAPENSVMTPPPVHYLOFKEMSDLSKYSPPPPQPPPELYVAASKHFQQAAMILESIPNADREVSRILKVAKPNEFVVMKLLAGGHKESKVPPEFDSVHKYFPVVKLV"

ORIGIN

Query Match 58.1%; Score 1799.2; DB 10; Length 2543;

Best Local Similarity 84.0%; Pred. No. 0;

Matches 2139; Conservative 0; Mismatches 343; Indels 64; Gaps 7;

QY 550 GGCATAATGGTTATGAAAGCTTCTCTAGATGATGACGATTCAGGATGGAGCTCAGTATG 609

Db 58 GGCATAATGGTTATGAAAGCCGCGAGTAGATGATGACGCTTCGGGATGGAGCTCAATGTC 117

QY 610 CCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAGATTTTGAAGAA 669

Db 118 CCTGAGAAAAATGGAAAAAAGTAGCACCAAGCTGGGTGGACATAACCCAGACTTTGAAGAT 177

QY 670 GCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGGTCTTTTGA 729

Db 178 GCTTGTGAGAGCTGAAAGTTGGGAGAACTGCTTCACGATAAGCTGTTGGTCTGTTGAA 237

QY 730 GCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTTGGAACCAA 789

Db 238 GCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCCGGGATGATCGGGAACCAA 297

QY 790 GTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAATAAA 849

Db 298 GTGAATAGAAAAAGTCCTCAATTTTGAACAAGCCGTCAAGGATGGCACCATTAAAAATAAA 357

QY 850 GATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTTTGTGCTTTTGATAACG 909

Db 358 GACCTCAGCTTGCTGAACTGATAGGAATAATGGATACCTGTTTTTGTGCTTTGATCACA 417

QY 910 TGGTTAGAAAGCCATTCACTGGCACAGACAGTATTTACGTGCCCTTTACATTCATAATCCA 969

Db 418 TGGCTCGAAGGCCATTCTTGGCACAGACGGTATTTACATGCTTTTACATTCATAATCCA 477

QY 970 GACTTTATAGAAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTGAAAAATCTGTGAC 1029

Db 478 GACTTCATAGAAAGACCCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTTGAAGATCTGTGAC 537

QY 1030 ATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGCTTTTTTTGAAGAGGAAGATTTTCAGTCAATG 1089

Db 538 ATTGCAAGGAAAAAGTCAACAAAGCTGCTGCTTTTGAGGAGGAAGATTTCAGTCAATG 597  
QY 1090 ACTTATGGATTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAA 1149  
Db 598 ACATATGGATTAAATGGCCAACGGGTGACAGATCTTCGAGTTACAGGGATGCTCAA 657  
QY 1150 GATGTGGAGGATGACATGCAAGAGAGTAAGAGTACTCGAAGTCGACAAAGGAGAAGAA 1209  
Db 658 GATGTGGAGGACGACATGCAAGAGGAGTAAGAGTACTCGAAGTCGACAAAGGAGAAG 717  
QY 1210 AGAGATCCAGAAAGTTGAAC TAGAACACCAACCAATGTTTAGCAGTATTCAGCAGAGTGAAA 1269  
Db 718 AGAGACCCAGAAGTTGAAC TAGAACACCAAGCAGTGTGGCAGCGTTTCAGCAGAGTGAAG 777  
QY 1270 TTTACTCGTGTGTACTGACAGTGCCTTATAGCCTTTACTAAGAAAGAGACCAAGTGT 1329  
Db 778 TTCACCCGAGTACTGCTGACAGTGTCTATAGCCTTCACTAAGAAAGAGACCAAGTGT 837  
QY 1330 GCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCANTCATAAATCA 1389  
Db 838 GCAGAGGCTCAGAAACTGATGGTCCAGGCGCAGACCTTCTTTCTGCCANTCACACCTCA 897  
QY 1390 TTGCATCATGGCATCCAGGCCCAGAAATGATFACATACAAAAGGAGATCATCCAAATTATGATG 1449  
Db 898 TTGCACACGGCATCCAGGCTCAGAAATGGCAATGGCACTACCAAAAGGAGACCATTCAAATTATGATG 957  
QY 1450 GGTTTTGAACCCCTGTGTGAAC CAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAA 1509  
Db 958 GGTTTTGAACCCCTGTGTGAAC CAGAGATTACTTCCACCCACCTTCCCTCGTATGCAAAA 1017  
QY 1510 ATAATTAAAGGGAAGAAATGGTGAACTATTTTGCAAGATTAAATAGATAGAAATAAAACT 1569  
Db 1018 ATAATTAAAGAGAAGAAATGGTCAACTATTTCTCAAGATTAAATAGACAGGATAAAACT 1077  
QY 1570 GTCTGTGAGGTGTGAATTTAACAAATTTACATTTGATCTCTGGATTTTCTGTGAATTT 1629  
Db 1078 GTCTGTGAAGTCGTGAACCTTACCAACTTTACATTTGATCTCTGGATTTTCTGTGAATTT 1137  
QY 1630 AGTGAACAGTCACCATGTGTTCTTTCAAGATCTGTGTTACAAACCACTTTCTCGTGGAT 1689  
Db 1138 AGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACTACTTTCTCGTGGAT 1197  
QY 1690 AACAAAAAGTCTTTGGAAC TCACTCAATGCAAGACATGGTGAAAGATGCACCTTCGGTCT 1749  
Db 1198 AACAAAAAGTCTTTGGGACTCATCTCATGCAAGACATGGTGAAAGATGCTCCCCGGTCC 1257  
QY 1750 TTTGTCAGATCCTCCGAGTGTCTTTCCCCCAAGTGTACCTATATAATAATCACCAGGCTA 1809  
Db 1258 TTCGTGAG--TCCTCCGGTGCTCTCCCCCAAGTGTGCCCTGTATAATAATCACCAGGCTA 1315  
QY 1810 AGGACTGTATCGACTCCTTTGTTACTCACTGTGTTGCGGCCATCTCTAGTCTTATTCAGA 1869  
Db 1316 AGGACTGCATCGACTCCTTTGTTACTCACTGTGTTGCGGCCATTTGTAGTCTTGTTCAGA 1375  
QY 1870 TCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATCTCTTGAGGAATTTG 1929  
Db 1376 TCCATGGCCATAACAGGGCTCGGCAGAGAGACAAACTTGGTCACATCTTGAAGAGTTTG 1435  
QY 1930 CCACCTTGACGGATGAGTTTAGACATTTTATTTAATAGGCAGAGAAGGTTGATGCAGC 1989  
Db 1436 CCACCTTGCAAGATG-----AGGCAGAGAAGGTCGATGCAGC 1472  
QY 1990 GCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGTAC 2049  
Db 1473 GCTCCACACTATGTTGTTGAAGCAGGAGCGCTCAGCGACAGCATCTGSCCTGTTTAGGAAC 1532  
QY 2050 CTGGTCCCTTTACCATAAACCTTCGCATTTATGATACAGTACCTTCTAAGTGGCTTTGAATT 2109  
Db 1533 TTGGGTTCTCTACCACAGCCTCGCGAATTATGATCCAGTACCTGCTCAGCGGCTTTGATCT 1592  
QY 2110 GGAAC TCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAAATTCCTTTA 2169

Db 1593 GGAGCTCTACAGCATGCACGAGTACTACTACATCTACTGGTACCTCTCTGAGTTCCTGTGA 1652  
QY 2170 CGCATGGTTGATGTCAACATTTGAGTCGTGCCGATGGCTCTCAAAATGGCAGAGAAAGGAT 2229  
Db 1653 CGCATGGCTGATGTCAACCTTGAGCCGTGCTGACGGCTCTCAGATGGCAGAAAGAGGAT 1712  
QY 2230 AATGGAAGAGCAGCAGAAAAGGCCGTAGTAGTAAAAAAAACAAAAGAAAAAAGAAAGTTTCG 2289  
Db 1713 AATGGAAGAAACAGCAGAAAAGGCCGCAGCAGCAAGAAAAACAAAGAAAGAAAGAAAGTTTCG 1772  
QY 2290 CCCATTGAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACATGTGTGCTGGAATGTT 2349  
Db 1773 TCCATTGAGCCGAGAGATCACAATGAGCCAGGCCCTATCAGAACATGTGTGCTGGATGTT 1832  
QY 2350 TAAAAACCATGTGTAGCATTTTGACATGGACGGCAAAAGTACGTAACCCGAAAGTTTGAGCTTGA 2409  
Db 1833 CAAAACGATGGTCGATTTTGACATGGATGGCAAAAGTGGCAAGCCCAAAATTTGAGCTTGA 1892  
QY 2410 TAGTGAACAAAGTTCGGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCC 2469  
Db 1893 CAGTGAGCAGGTCGGTATGAGCACAGGTTTGCTCCGTTCAACAGCGTGTGATGACGCCACC 1952  
QY 2470 GCCAGTGCATCTACTTACAGTTCAAGGAATGTCTGACCTCAATAAATATAGCCCTCCTCC 2529  
Db 1953 ACCAGTGCATCTACTGACGTTCAAGGAATGTCTGACCTCAGTAAATACAGCCCTCCTCC 2012  
QY 2530 TCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTTCAACAGGCAAAAAATGATATT 2589  
Db 2013 TCAGCCTCCAGAGCTGTATGTGGCAGCTAGTAAGCACTTTTCAGCAGGCAAAAGATGATCCT 2072  
QY 2590 GGAATAATATTCCTAACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCAA 2649  
Db 2073 TGAGAGCATCCCCAATGCAGACCCGGGAGGTCAGCAGGATCCTTAAAGGTTGCCAAGCCTAA 2132  
QY 2650 CTTTGTGGTTATGAAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCTCTCCTGA 2709  
Db 2133 CTTTGTGGTTATGAAAGCTCTTAGCAGGAGGACACAAAGAGTCAAAGGTTCTCTCCTGA 2192  
QY 2710 ATTTGATTTCTCTGCTCATAAATATTTTCTCTGTGTGAAAACCTTGTGTGAGAGACTGGG 2769  
Db 2193 ATTTGATTTCTCTGTTCACAAATACTTTCCTGTGTGAAAACCTGTTTGAAGAGACTGCA 2252  
QY 2770 GAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCAACTCTTAGAGGGGCACATCAC 2829  
Db 2253 AAGATGACCAT-----GTCTACTCTGGGGTACACCA 2283  
QY 2830 AGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACTCATATAAGGA--- 2886  
Db 2284 GTGTGCCACACACGGTGACATAAAGTGGCTCTCTTGAATGACAGCTGTTTACAGGAAGAC 2343  
QY 2887 -ATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACCTGCTGTTTAAAGTGGTT 2945  
Db 2344 TTCAGTATCAGTACAACACCCCTTACCTGGAGAGAAG- AAAC TGTGTTTAAAGTGGTT 2402  
QY 2946 TATTATGTTCCATGGAAGAAAC TGGTCTTATTGAATGCATTTGTAACGTTATATGTTT 3005  
Db 2403 TGTATACTCCATGGGTGTGACGGG----CTGGAATGCATTTGGTGAACGTTTACATGGTTT 2458  
QY 3006 TATTACAGATTTAATCAAAAATCAATTTTTTATGAATGATGAGTGAATAAGTGTATTATA 3065  
Db 2459 TATTACAGACTTCATTTGTAATACTTTTAAATGAATGATTTGTGTG-AAATAGTGTTTGTGA 2517  
QY 3066 AAGGTTAATAAATTTCTTGACAAAAA 3091  
Db 2518 AAGGTTAATAAATTTCTTGACAAAAA 2543

RESULT 6  
AK025266  
LOCUS AK025266 1839 bp mRNA linear PRI 13-SEP-2003  
DEFINITION Homo sapiens cDNA: FLJ21613 fis, clone COL07381.  
ACCESSION AK025266  
VERSION AK025266.1 GI:10437744



KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1839)

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission

JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source Location/Qualifiers

1. .1839

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="COL07381"

/tissue type="colon"

/clone\_lib="COL"

/note="cloning vector pME18SFL3"

134. .1714

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB15097.1"

/db\_xref="GI:10437745"

/translation="MVMKASVDDDDSGWELSMPEKMEKSNWNVDTQDFEEACREBLK LGELLHDKLFGLFEMSALEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTL PELIGIMDTFCCLITWLEGHSIAQTVFTCLYIHNPDFIEDPAMKAFALGILKICDIA REKVNKAAVFEEDFQSMYGFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEE RDPEVELEHQRCIAVFSRVKFTRVLLTVLIAFTKETSAVAEAAQKLMVQAADLLSAIH NSLHHGIOAQNDDTKGDHPIMMGFEPLVNQRLLPPTFPYAKIIKREEMVNYFARLID RIKTVCEVNVNLNLHCILDFCEFESEQSPCVLSRSILQTTFLVDNKKVFQFTHLMODMV KDALRSFVPPVLSPKVLYNNHQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKL GHILEEFATLQDEAEKVDAAALHTMLLKQEPQRQHLAWLGTWVLYHNLRIMIQYLLSGF EELYSMHEILLHILVSL"

CDS

ORIGIN

Query Match 57.1%; Score 1766.6; DB 9; Length 1839;

Best Local Similarity 98.1%; Pred. No. 0;

Matches 1828; Conservative 0; Mismatches 9; Indels 26; Gaps 3;

QY 423 ATACGCATGCGTGACCGTGCCGGTGGGCTGGGCTGAGAGGGGAGGGCGGCGCGGC 482

DB 1 ATACGCATGCGTGACCGTGCCGGTGGGCTGGGCTGAGAGGGGAGGGCGGCGCGGC 60

QY 483 CGAGGCGGCGTCGTTATTTCCGTGGTCCGACAGTGCCTGGCGGCGCGGCGGACCCACGGG 542

DB 61 CGAGGCGGCGTCGTTATTTCCGTGGTCCGACAGTGCCTGGCGGCGGCGGCGGACCCACGGG 120

QY 543 AGAAGTAGGCAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCT 602

DB 121 AGAAGTAGGCAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCT 180

QY 603 CAGTATGCCAGAAAAAATGGAGAAAAAGCAATAACAAACTGGGTGGACATACCCAAAGATT 662

DB 181 CAGTATGCCAGAAAAAATGGAGAAAAAGCAATAACAAACTGGGTGGACATACCCAAAGATT 240

QY 663 TGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCT 722

DB 241 TGAAGAAGCTTGTGAGAAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCT 300

QY 723 TTTTGAAGCCCATGTCTCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGG 782

DB 301 TTTTGAAGCCCATGTCTCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGG 360

QY 783 AAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGCTATTAA 842

DB 361 AAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGCTATTAA 420

QY 843 AATTAAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTTTGGCTGTT 902

DB 421 AATTAAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTTTTGGCTGTT 480

QY 903 GATTAACGTGGTTAGAAGGCCAATTCACCTGGCACAGACAGTANTTACGTCCCTTTACATTCA 962

DB 481 GATAACGTGGTTAGAAGGCCAATTCACCTGGCACAGACAGTANTTACGTCCCTTTACATTCA 540

QY 963 TAATCCAGACTTTTATAGAAGATCTGCTATGAAGGCTTTTGGCTCTGGGAATCTTGAAAAT 1022

DB 541 TAATCCAGACTTTTATAGAAGATCTGCTATGAAGGCTTTTGGCTCTGGGAATCTTGAAAAT 600

QY 1023 CTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCA 1082

DB 601 CTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCA 660

QY 1083 GTCAATGACTTATGGATTTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCAT 1142

DB 661 GTCAATGACTTATGGATTTTAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCAT 720

QY 1143 GCTAAAAGATGTGGAGGATGACATGCAAAAGAAAGAGTAAAGAGTACTCGAAGTCGACAAG 1202

DB 721 GCTAAAAGATGTGGAGGATGACATGCAAAAGAAAGAGTAAAGAGTACTCGAAGTCGACAAG 780

QY 1203 AGAAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTACAGCAG 1262

DB 781 AGAAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAAGATGTTTAGCAGTATTACAGCAG 840

QY 1263 AGTGAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAGAAAGAGACCAG 1322

DB 841 AGTGAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCTTTTACTAAGAAAGAGACCAG 900

QY 1323 TGCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCA 1382

DB 901 TGCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCA 960

QY 1383 TAATTCATTGCATCATGGCATCCAGGCCAGAAATGATATCTACAAAAGGAGATCATCCAAT 1442

DB 961 TAATTCATTGCATCATGGCATCCAGGCCAGAAATGATATCTACAAAAGGAGATCATCCAAT 1020

QY 1443 TATGATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATA 1502

DB 1021 TATGATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATA 1080

QY 1503 TGCAAAAATAATTAAAAGGGAGAAATGGTGAACCTATTTTGCAGATTAATAGATAGAAT 1562

DB 1081 TGCAAAAATAATTAAAAGGGAGAAATGGTGAACCTATTTTGCAGATTAATAGATAGAAT 1140

QY 1563 AAAAACTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATTCCTGGATTTTTCTG 1622

DB 1141 AAAAACTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATTCCTGGATTTTTCTG 1200

QY 1623 TGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCTCT 1682

DB 1201 TGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCTCT 1260

QY 1683 GGTGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCAC 1742

DB 1261 GGTGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCAC 1320

QY 1743 TCGGTCTTTTGTGATCATCCTCCGAGTGTCTTTCCCCCAAGTGTCTACCTATATAATATCAC 1802







RESULT 8  
AK026296  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AK026296  
Homo sapiens cDNA: FLJ22643 fis, clone HSI07031.  
AK026296  
AK026296.1 GI:10439112  
oligo capping; fis (full insert sequence).  
Homo sapiens (human)

REFERENCE  
AUTHORS

Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,  
Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
Nakamura,Y., Isogai,T. and Sugano,S.  
NEDO human cDNA sequencing project

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 997)  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
Direct Submission  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Departent of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source

Location/Qualifiers  
1. .997  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HSI07031"  
/tissue\_type="human small intestine"  
/clone\_lib="HSI"  
/note="cloning vector pME18SFL3"  
15. .650  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15435.1"  
/db\_xref="GI:10439113"  
/translation="MHEYVYIYWLYSEFLYAWLMSILSRADGSQMAEERIMEEQKGR  
SSKTKTKKKVRPLSREITMSOAYQNMCAGMFKTMVAFDMGKVRKPKFELDSQVRY  
EHRFPAPFNSVMTPPVHYLPFKEMSDLNKYSPPQSPPELYVAASKHFQQAQKMLENIP  
NPDHEVNRILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSAHKYFPVVKLV"

CDS

ORIGIN

Query Match 31.9%; Score 988; DB 9; Length 997;  
Best Local Similarity 100.0%; Pred. No. 3.4e-213;  
Matches 988; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2109 TGGAACTTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 2168  
|||||  
Db 1 TGGAACTTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTT 60  
|||||

Qy 2169 ACGCATGTTGATGTCAACATTGAGTCGTCGCGATGGCTCTCAAAATGGCAGAGGAAAGGA 2228  
|||||  
Db 61 ACGCATGTTGATGTCAACATTGAGTCGTCGCGATGGCTCTCAAAATGGCAGAGGAAAGGA 120  
|||||

Qy 2229 TAATGGAAGCAGCAGAGAAAGCCGCTAGTAGTAAACAAAGAAAAAGAAAGTTTC 2288  
|||||  
Db 121 TAATGGAAGCAGCAGAGAAAGCCGCTAGTAGTAAACAAAGAAAAAGAAAGTTTC 180  
|||||

Qy 2289 GCCCATTTGACCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAATGT 2348  
|||||  
Db 181 GCCCATTTGACCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGAATGT 240  
|||||

QY 2349 TTAAACCATGGTAGCATTTGACATGGACGGCAAAGTACGTAAACCGAAGTTTGAGCTTG 2408  
|||||  
Db 241 TTAAACCATGGTAGCATTTGACATGGACGGCAAAGTACGTAAACCGAAGTTTGAGCTTG 300  
|||||

QY 2409 ATAGTGAACAAAGTTCGGTATGAACACAGGTTTGCTCCATTCAAAGTGTGATGACCCCGC 2468  
|||||  
Db 301 ATAGTGAACAAAGTTCGGTATGAACACAGGTTTGCTCCATTCAAAGTGTGATGACCCCGC 360  
|||||

QY 2469 CGCCAGTGCACACTTACAGTTCAAGGAAATGTCTGACCTCAAPAAATATAGCCCTCCTC 2528  
|||||  
Db 361 CGCCAGTGCACACTTACAGTTCAAGGAAATGTCTGACCTCAAPAAATATAGCCCTCCTC 420  
|||||

QY 2529 CTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAGCACCTTCAACAGGCAAAAATGATAT 2588  
|||||  
Db 421 CTCAGTCTCTGAACTGTATGTGGCAGCTAGTAAGCACCTTCAACAGGCAAAAATGATAT 480  
|||||

QY 2589 TGGAAAATATTCTTAACCCGGACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCA 2648  
|||||  
Db 481 TGGAAAATATTCTTAACCCGGACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCA 540  
|||||

QY 2649 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCCTG 2708  
|||||  
Db 541 ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCCTG 600  
|||||

QY 2709 AATTGCAATTCCTGCTCATAAATATTTTCTGTGTGAAACTTGTGTGAGAGAGACTGG 2768  
|||||  
Db 601 AATTGCAATTCCTGCTCATAAATATTTTCTGTGTGAAACTTGTGTGAGAGAGACTGG 660  
|||||

QY 2769 GGAGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCACTCTTAGAGGGCACATCAC 2828  
|||||  
Db 661 GGAGTGGCCATAAAGGGGCAGAGTCTTCTTTTCAGACCCCACTCTTAGAGGGCACATCAC 720  
|||||

QY 2829 CAGGCTCCACATCACCGGAAGTCGAGATGGATTCTTGGGTAACAACTCATTATAAGGAAT 2888  
|||||  
Db 721 CAGGCTCCACATCACCGGAAGTCGAGATGGATTCTTCTGGGTAACAACTCATTATAAGGAAT 780  
|||||

QY 2889 ACTTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAGTGGTTTAT 2948  
|||||  
Db 781 ACTTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAGTGGTTTAT 840  
|||||

QY 2949 TATGTTCCATGGAAGAAACTGGTCTTATTGAATGCAATGATGAACGTTATATGTTTAT 3008  
|||||  
Db 841 TATGTTCCATGGAAGAAACTGGTCTTATTGAATGCAATGATGAACGTTATATGTTTAT 900  
|||||

QY 3009 TACAGATTTAATCACAATCATTTTTTATGAATGATTGAGTGAATAATAGTGTATAAAG 3068  
|||||  
Db 901 TACAGATTTAATCACAATCATTTTTTATGAATGATTGAGTGAATAATAGTGTATAAAG 960  
|||||

QY 3069 GTTAATAAATTTCTTGACAAAAAAGAAAA 3096  
|||||  
Db 961 GTTAATAAATTTCTTGACAAAAAAGAAAA 988  
|||||

RESULT 9  
AK056059  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AK056059  
Homo sapiens cDNA FLJ31497 fis, clone NT2NE2005317, highly similar  
to Rattus norvegicus corneal wound healing related protein mRNA.  
AK056059  
AK056059.1 GI:16551204  
oligo capping; fis (full insert sequence).  
Homo sapiens (human)  
Homo sapiens

REFERENCE  
AUTHORS

1  
Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,  
Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,  
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,  
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,  
Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,



Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 1796)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES  
source  
1. .1796  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2NE2005317"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2NE2"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuron after the differentiation of NT2 neuronal precursor cells."

ORIGIN

Query Match 28.6%; Score 885.4; DB 9; Length 1796;  
Best Local Similarity 99.9%; Pred. No. 7.2e-190;  
Matches 886; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 549 AGGCAPAAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAGTAT 608  
Db |||||||  
QY 345 AGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAGTAT 404  
Db |||||||  
QY 609 GCCAGAAAAAATGGAGAAAAGCAATACAAAACCTGGGTGGACATTACCCAAAGATTTTGAAGA 668  
Db |||||||  
QY 405 GCCAGAAAAAATGGAGAAAAGCAATACAAAACCTGGGTGGACATTACCCAAAGATTTTGAAGA 464  
QY 669 AGCTTGTCTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTGGTCTTTTGA 728  
Db |||||||  
QY 465 AGCTTGTCTGAGAAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTGGTCTTTTGA 524  
QY 729 AGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCA 788  
Db |||||||  
QY 525 AGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCA 584  
QY 789 AGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTA 848  
Db |||||||  
QY 585 AGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTA 644  
QY 849 AGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTGTGCTGTTTGATAAC 908  
Db |||||||  
QY 645 AGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTGTGCTGTTTGATAAC 704  
QY 909 GTGGTTAGAAGGCCATTCACTGGCAGACAGCATTTTACGTGCCTTTACATTCATAATCC 968  
Db |||||||  
QY 705 GTGGTTAGAAGGCCATTCACTGGCAGACAGCATTTTACGTGCCTTTACATTCATAATCC 764  
QY 969 AGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTGAAAATCTGTGA 1028  
Db |||||||  
QY 765 AGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTCTGGGAATCTTGAAAATCTGTGA 824  
QY 1029 CATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAAT 1088  
Db |||||||  
QY 825 CATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTCAAT 884  
QY 1089 GACTTATGGATTTAAATGGCTTAACAGTGTGACAGATCTTTCGAGTTACAGGCATGCCTAAA 1148

Db 885 GACTTATGGATTTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAA 944  
QY 1149 AGATGTGGAGGATGACATGCAAAGAAGAGTAAAGAGTACTCGAAGTTCGACAAGGAGAAGA 1208  
Db |||||||  
QY 945 AGATGTGGAGGATGACATGCAAAGAAGAGTAAAGAGTACTCGAAGTTCGACAAGGAGAAGA 1004  
QY 1209 AAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTTCAGCAGAGTGAA 1268  
Db |||||||  
QY 1005 AAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTTCAGCAGAGTGAA 1064  
QY 1269 ATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACCAGTGTGT 1328  
Db |||||||  
QY 1065 ATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACCAGTGTGT 1124  
QY 1329 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATAATTC 1388  
Db |||||||  
QY 1125 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATAATTC 1184  
QY 1389 ATTGCATCATGGCATCCAGGCCCCAGAATGATACTACAAAAGGAGATC 1435  
Db |||||||  
QY 1185 ATTGCATCATGGCATCCAGGCCCCAGAATGATACTACAAAAGGAGGTC 1231

RESULT 10  
LOCUS AX664135 719 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 7 from Patent WO02057304.  
ACCESSION AX664135  
VERSION AX664135.1 GI:29164145  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Panzer,S.R., Lincoln,S.E., Altus,C.M., Dufour,G.E., Jackson,J.L., Jones,A.L., Dam,T.C., Liu,T.F., Harris,B., Flores,V., Daffo,A., Marwaha,R., Chen,A.J., Chang,S.C., Gerstin,E.H., Peralta,C.H., David,M.H. and Lewis,S.A.  
TITLE Secretory molecules  
JOURNAL Patent: WO 02057304-A 7 25-JUL-2002;  
Incyte Genomics, Inc. (US)

FEATURES  
source  
1. .719  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: LI:462889.1:2001JAN12"

ORIGIN

Query Match 23.2%; Score 717.4; DB 6; Length 719;  
Best Local Similarity 99.9%; Pred. No. 8.4e-152;  
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 474 GCGCGCGCGGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTGC GTGGCGCGCGGT 533  
Db |||||||  
QY 1 GCGCGCGCGGAGCGCGCTCGTTATTTCCGTGGTCCGGACAGTGC GTGGCGCGCGGT 60  
QY 534 GACCACGGGAGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593  
Db |||||||  
QY 61 GACCACGGGAGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120  
QY 594 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAACTGGGTGGACATTAC 653  
Db |||||||  
QY 121 ATGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAACTGGGTGGACATTAC 180  
QY 654 CCAAGATTTTGAAGAAGCTTTCGAGAAATTAAAGTTGGGAAACTACTTTCATGATAAGCT 713  
Db |||||||  
QY 181 CCAAGATTTTGAAGAAGCTTTCGAGAAATTAAAGTTGGGAAACTACTTTCATGATAAGCT 240  
QY 714 ATTTGGTCTTTTGAAGCCATGCTCTGCTATTGAAATGATGGATCCCAGATGGATGCTGG 773

Db 241 ATTTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 300

QY 774 CATGATTGGAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833

Db 301 CATGATTGGAACCAAGTTAATCGAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360

QY 834 CACTATTAAATTAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTT 893

Db 361 CACTATTAAATTAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTT 420

QY 894 TTGCTGTTTGTATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTACGTGCCT 953

Db 421 TTGCTGTTTGTATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTACGTGCCT 480

QY 954 TTACATTCAATCCAGACTTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTTGGGAAT 1013

Db 481 TTACATTCAATCCAGACTTTTATAGAGATCCTGCTATGAAGGCTTTTGTCTTGGGAAT 540

QY 1014 CTTGAAATCTGTGACATTGCAAGGGGAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGA 1073

Db 541 CTTGAAATCTGTGACATTGCAAGGGGAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGA 600

QY 1074 AGATTTTCAGTCAATGACTTATGGATTTAAATGGCTTAACAGTGTGACAGATCTTCGAGT 1133

Db 601 AGATTTTCAGTCAATGACTTATGGATTTAAATGGCTTAACAGTGTGACAGATCTTCGAGT 660

QY 1134 TACAGGCATGCTAAAGATGTGGAGGATGACATGCAAAGAGAGTAAAGAGTACTCGAA 1192

Db 661 TACAGGCATGCTAAAGATGTGGGGATGACATGCAAAGAGAGTAAAGAGTACTCGAA 719

RESULT 11

AC102010

LOCUS AC102010 147876 bp DNA linear HTG 21-MAR-2003

DEFINITION Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered pieces.

ACCESSION AC102010

VERSION AC102010.3 GI:29135693

KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 147876)

Birren,B., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 147876)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL

REFERENCE

AUTHORS

3 (bases 1 to 147876)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 21, 2003 this sequence version replaced gi:28412092.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

TITLE

JOURNAL

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L17873

Center clone name: 531\_M\_24

----- Summary Statistics

COMMENT

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 146099 bases at least Q40

Consensus quality: 146869 bases at least Q30

Consensus quality: 147004 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 147276; sum-of-contigs

Quality coverage: 9.9 in Q20 bases; agarose-fp

Quality coverage: 10.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 7 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 84958: contig of 84958 bp in length

\* 84959 85058: gap of 100 bp

\* 85059 87411: contig of 2353 bp in length

\* 87412 87511: gap of 100 bp

\* 87512 89933: contig of 2422 bp in length

\* 89934 90033: gap of 100 bp

\* 90034 94076: contig of 4043 bp in length

\* 94077 94176: gap of 100 bp

\* 94177 102536: contig of 8360 bp in length

\* 102537 102636: gap of 100 bp

\* 102637 118954: contig of 16318 bp in length

\* 118955 119054: gap of 100 bp

\* 119055 147876: contig of 28822 bp in length.

FEATURES

source

1.147876

/organism="Mus musculus"

/mol\_type="genomic DNA"





TITLE  
JOURNAL  
  
COMMENT  
  
Direct Submission  
Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
On Sep 11, 2001 this sequence version replaced gi:14267784.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em.: EMBL; Sw.:  
SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-213G2 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6  
IMPORTANT: This sequence is not the entire insert of clone  
RP11-213G2 It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true right end of clone RP11-213G2 is at 181864 in this  
sequence. The true left end of clone RP11-65C15 is at 111320 in  
this sequence. The true right end of clone RP11-202I11 is at 100 in  
this sequence.  
FEATURES  
source  
1..181864  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-213G2"  
/clone\_lib="RPCI-11.1"  
39727  
misc\_feature  
115957..116075  
/note="Sequence from overlapping clone BA65C15 (AL161453).  
Assembly confirmed by restriction digest."  
117488..117564  
/note="Sequence from overlapping clone BA65C15 (AL161453).  
Assembly confirmed by restriction digest."  
ORIGIN  
Query Match 17.8%; Score 551.4; DB 9; Length 181864;  
Best Local Similarity 99.8%; Pred. No. 9.4e-114;  
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TTCCTCAGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTGTCTCAGGCAGCAGTA 60  
Db 174760 TTCCTCAGGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTGTCTCAGGCAGCAGTA 174819  
QY 61 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTCGCTCGCAACCACTAAGGTCTACG 120  
Db 174820 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTCGCTCGCAACCACTAAGGTCTACG 174879  
QY 121 CAAACCTCCACGGTTTCCTTCGGCCCTTCGGGTACCTTTCTAAGAAATTTCCAGAGGGCA 180  
Db 174880 CAAACCTCCACGGTTTCCTTCGGCCCTTCGGGTACCTTTCTAAGAAATTTCCAGAGGGCA 174939  
QY 181 GCGCAGCGGGCGGGCTCTGAGACTCCGGGCTCCGCCCTCTTTCCGGGAACCGCCCACTA 240

Db 174940 GCGCAGACGGGGCGGGCTCTGAGACTCCGGGGCTCCGCCTCTTTCCGGGAACCGCCCACTA 174999  
QY 241 CCCAGGACTCCGACAGAGGGTGAAAAAAGATAAACTTCCGGTCTCGCGATCGTCTCTAATC 300  
Db 175000 CCCAGGACTCCGACAGAGGGTGAAAAAAGATAAACTTCCGGTCTCGCGATCGTCTCTAATC 175059  
QY 301 TCGCGAGAAGAGAGGGCGGCCCATCGGCCGAAACGAGGCGGTGGCGAGGGGGGTG 360  
Db 175060 TCGCGAGAAGAGAGGGCGGCCCATCGGCCGAAACGAGGCGGTGGCGAGGGGGGTG 175119  
QY 361 TGGCCGGGAGCGCGAAGTCCCCGGGAGTAAGGAGAGGGGGCGGGGTCCGCCGTCCCGG 420  
Db 175120 TGGCCGGGAGCGCGAAGTCCCCGGGAGTAAGGAGAGGGGGCGGGGTCCGCCGTCCCGG 175179  
QY 421 GCATACGCATCGGTGCACGCTGCCGTCGGGCTGGGCTGAGAGGGGAGGGGGCGGCG 480  
Db 175180 GCATACGCATCGGTGCACGCTGCCGTCGGGCTGGGCTGAGAGGGGAGGGGGCGGCG 175239  
QY 481 GCCGAGGCGGCGTCTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGGGGTGACCAAG 540  
Db 175240 GCCGAGGCGGCGTCTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGGGGTGACCAAG 175299  
QY 541 GGAGAAAGTAGGCA 553  
Db 175300 GGAGAAAGTAGGTA 175312  
  
RESULT 13  
AX778899  
LOCUS AX778899 440 bp DNA linear PAT 14-JUL-2003  
DEFINITION Sequence 1056 from Patent WO03039443.  
ACCESSION AX778899  
VERSION AX778899.1 GI:32695893  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Haerlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,  
Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.  
TITLE Novel genetic markers for leukemias  
JOURNAL Patent: WO 03039443-A 1056 15-MAY-2003;  
Deutsches Krebsforschungszentrum (DE) ;  
Ludwig-Maximilian-Universitaet Muenchen (DE) ;  
PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)  
FEATURES  
source  
1..440  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 14.2%; Score 440; DB 6; Length 440;  
Best Local Similarity 100.0%; Pred. No. 7e-89;  
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2548 TGTGGCAGCTAGTAAGCACCTTCAACAGGCAGGCAAAATGATATTGAAAATATTCCTAACCC 2607  
Db 1 TGTGGCAGCTAGTAAGCACCTTCAACAGGCAGGCAAAATGATATTGAAAATATTCCTAACCC 60  
QY 2608 GGACCATGAGTTAATAGAAATTTAAAGGTTGCCAAACCCAACTTTGTGGTTATGAAGTT 2667  
Db 61 GGACCATGAGTTAATAGAAATTTAAAGGTTGCCAAACCCAACTTTGTGGTTATGAAGTT 120  
QY 2668 ATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCTGAATTTGATTTCTCTGCTCA 2727  
Db 121 ATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCTGAATTTGATTTCTCTGCTCA 180  
QY 2728 TAAATATTTTCTGTGAAACTTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGG 2787  
Db 181 TAAATATTTTCTGTGAAACTTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGG 240



QY 2788 CAGAGTCTTCTTTTCAGACCCCACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGGA 2847  
|||||  
Db 241 CAGAGTCTTCTTTTCAGACCCCACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGGA 300  
|||||  
QY 2848 AGTCAGATGGATTTCTTTGGGTAAACAACCTCATTATAAGGAATACTTTTAGTTTGACGCCT 2907  
|||||  
Db 301 AGTCAGATGGATTTCTTTGGGTAAACAACCTCATTATAAGGAATACTTTTAGTTTGACGCCT 360  
|||||  
QY 2908 TATATGACATGAATGAAAACTGCTGTTTTTAAAGTGTTTATATGTTCCATGGAAGAAAC 2967  
|||||  
Db 361 TATATGACATGAATGAAAACTGCTGTTTTTAAAGTGTTTATATGTTCCATGGAAGAAAC 420  
|||||  
QY 2968 TGGTCTTATTGAATGCATTG 2987  
|||||  
Db 421 TGGTCTTATTGAATGCATTG 440  
|||||

RESULT 14  
AL161447  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-379P1 on chromosome 9, complete sequence.  
AL161447  
ACCESSION AL161447.20 GI:19572746  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 160115)  
Johnson,C.  
Direct Submission  
Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk humquery@sanger.ac.uk  
On Mar 21, 2002 this sequence version replaced gi:15787718.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-379P1 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-379P1 The true left end of clone RP11-280P22 is at 137145 in this sequence. The true right end of clone RP11-65C15 is at 38497 in this sequence.  
Location/Qualifiers  
1. 160115  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-379P1"  
/clone\_lib="RPCI-11.2"

FEATURES  
source

misc\_feature complement (32505..32575)  
/note="Sequence from uni-directional dGTP big dye terminator reads only."  
74372..74537  
misc\_feature /note="Single clone region. Assembly confirmed by restriction digest data."  
74706..74920  
misc\_feature /note="Single clone region. Assembly confirmed by restriction digest data."  
complement (74706..74920)  
misc\_feature /note="Sequence from uni-directional dGTP big dye terminator reads only."  
74921  
misc\_feature /note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."  
129108  
misc\_feature /note="Tandem repeat. Forced join. Gap size estimated to be approximately 300bp by restriction digest data."

ORIGIN

Query Match 12.9%; Score 398.4; DB 9; Length 160115;  
Best Local Similarity 99.8%; Pred. No. 4.9e-79;  
Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2697 AAGTTCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTCTGTGTGAAACTTGTTT 2756  
|||||  
Db 43136 AGGTTCTCCTGAATTTGATTTCTCTGCTCATAAATATTTCTGTGTGAAACTTGTTT 43195  
|||||  
QY 2757 GAGAGAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCCTTTTCAGACCCCACTCTTAG 2816  
|||||  
Db 43196 GAGAGAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCCTTTTCAGACCCCACTCTTAG 43255  
|||||  
QY 2817 AGGGCACATCACCCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACCTC 2876  
|||||  
Db 43256 AGGGCACATCACCCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACCTC 43315  
|||||  
QY 2877 ATTATAAGGAATACATTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTT 2936  
|||||  
Db 43316 ATTATAAGGAATACATTTTAGTTTGACAGCCTTATATGACATGAATGAAACTGCTGTTT 43375  
|||||  
QY 2937 AAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATGAATGCATTGATGAACGTT 2996  
|||||  
Db 43376 AAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATGAATGCATTGATGAACGTT 43435  
|||||  
QY 2997 ATATGGTTTTTATTACAGATTTAATCACAAATCATTTTTTTATGAATGATTGAGTGAATA 3056  
|||||  
Db 43436 ATATGGTTTTTATTACAGATTTAATCACAAATCATTTTTTTATGAATGATTGAGTGAATA 43495  
|||||  
QY 3057 GTGTTTATAAAGGTTAATAAATTTCTTGACAAAAA 3096  
|||||  
Db 43496 GTGTTTATAAAGGTTAATAAATTTCTTGACAAAAA 43535  
|||||

RESULT 15  
HSA334821/c  
LOCUS  
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone HSA334821 446 bp DNA linear PRI 18-JUL-2002  
NB6-015R.  
ACCESSION AJ334821  
VERSION AJ334821.1 GI:15879239  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 446)  
REFERENCE Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M., Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V., Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.  
TITLE NotI flanking sequences: a tool for gene discovery and verification of the human genome  
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)

MEDLINE 22131767  
PUBMED 12136098  
REFERENCE 2 (bases 1 to 446)  
AUTHORS Zabarovsky,E.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,  
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,  
Sweden

FEATURES  
source Location/Qualifiers  
1..446  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="NB6-015R"

ORIGIN

Query Match 10.4%; Score 321.4; DB 9; Length 446;  
Best Local Similarity 99.7%; Pred. No. 5.8e-62;  
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCCTCAGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 60  
|||  
Db 323 TTCCTCAGGACACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 264  
|||

QY 61 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTCGTGCTCGCAACCACTAAGGTCTACG 120  
|||  
Db 263 GCACGCTGTGCTCTCGGAGCTTGGCTGCTCGTTCGTGCTCGCAACCACTAAGGTCTACG 204  
|||

QY 121 CAAACCTCCACGGTTTCCTTCCGCCCTTCGCGTCACCTTTCTAAGAAATCCAGAGGGCA 180  
|||  
Db 203 CAAACCTCCACGGTTTCCTTCCGCCCTTCGCGTCACCTTTCTAAGAAATCCAGAGGGCA 144  
|||

QY 181 GCGCAGACGGGGCGGGCTCTGAGACTCCGGCTCCGCCCTCTTCCGGGAACCGCCCACTA 240  
|||  
Db 143 GCGCAGACGGGGCGGGCTCTGAGACTCCGGCTCCGCCCTCTTCCGGGAACCGCCCACTA 84  
|||

QY 241 CCCAGGACTCCGACAGAGGGTGAAAAAGATAAATTCCGGTCTCGGATCGTCTCTAATC 300  
|||  
Db 83 CCCAGGACTCCGACAGAGGGTGAAAAAGATAAATTCCGGTCTCGGATCGTCTCTAATC 24  
|||

QY 301 TCGCGAGAAGAGAGGCGGCGC 323  
|||  
Db 23 TCGCGAGAAGAGAGGCGGCGC 1  
|||

Search completed: August 11, 2004, 01:34:09  
Job time : 12343.3 secs















JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM9999 row: 1 column: 07  
High quality sequence stop: 785.  
FEATURES Location/Qualifiers  
source 1..854  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4359702"  
/tissue\_type="adrenal cortex carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC 84"  
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 22.5%; Score 696.8; DB 10; Length 854;  
Best Local Similarity 94.5%; Pred. No. 7.5e-120;  
Matches 809; Conservative 0; Mismatches 37; Indels 10; Gaps 8;  
QY 2245 GAAAGGCCGTAGTAGTAAAAAACAAG-AAAAAAAAGAAAGTTCGCCCATTTGAGCCGAG 2303  
Db 1 GAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAAAGAAAGTTCGCCCATTTGAGCCGAG 60  
QY 2304 AGATCACAATGAGCCAAAGCATATCAGAACATGTGTGCTGGAATGTTTAAAAACCATGGTAG 2363  
Db 61 AGATCACAATGAGCCAAAGCATATCAGAACATGTGTGCTGGAATGTTTAAAAACCATGGTAG 120  
QY 2364 CATTTGACATGACGCGCAAGTACGTAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 2423  
Db 121 CATTTGACATGACGCGCAAGTACGTAACCGAAGTTTGAGCTTGATAGTGAACAAAGTTC 180  
QY 2424 GGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCGAGTGCACACTACT 2483  
Db 181 GGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCGAGTGCACACTACT 240  
QY 2484 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAAC 2543  
Db 241 TACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAAC 300  
QY 2544 TGTATGTGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATGGAAAAATATTCCTA 2603  
Db 301 TGTATGTGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATGGAAAAATATTCCTA 360  
QY 2604 ACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTGTGGTTATGA 2663  
Db 361 ACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTGTGGTTATGA 420  
QY 2664 AGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCCTCCTGAATTTGATTTCTCTG 2723  
Db 421 AGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCCTCCTGAATTTGATTTCTCTG 480  
QY 2724 CTCATAAATATTTTCCTGTTGTGAAACTTGTGAGAGAGACTGGGGAGGTGGCCATAAA 2783  
Db 481 CTCATAAATATTTTCCTGTTGTGAAACTTGTGAGAGAGACTGGGGAGGTGGCCATAAA 540  
QY 2784 GGGCAGAGTCTTCTTTCAGACCCCACTCTTAGAGGGCACATCACAGGCTCCACATCAC 2843  
Db 541 GGGCAGAGTCTTCTTTCAGACCCCACTCTTAGAGGGCACATCACAGGCTCCACATCAC 600  
QY 2844 GGGAAGTGAGATGGATTTCTTGGGTAACAACACTCATTATAAGGAATACTTTTAG-TTTGAC 2902

Db 601 GGGAAGTGAGATGGATTCTTGGGTAAACAACTCATTATAAGGAATACTTTAGTTTGAC 660  
QY 2903 AGCCTTATATGACA-TGAATGAAAACACTGCTGTTTAA-AGTGGTTTATATGTTCCAT-G 2959  
Db 661 AGCCTTATATGACATTGAATGAAAACACTGCTGTTTAAACAGTGGTCTTTATGTTCCCTGG 720  
QY 2960 GAAGAAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTTATACAGATTAA 3019  
Db 721 GACGAAACTGGTCTTATTGAATGGCTTGATGAACG-TATATGGGTTTATACCGATTTA- 778  
QY 3020 TCACAAATCATTTTTTATGAATGATTGAGTGAATAAGTGTGTTTATAAGGTTAATAAATT 3079  
Db 779 --TCCCAATCATTTTTTATGAATGATTG-GTGCACATCGGTTTTAACGTGATAAAATTCTTG 835  
QY 3080 TCCTTGACAAAAAAA 3095  
Db 836 CGCAAAAAAACAAAAAA 851  
RESULT 5  
LOCUS CB169246 814 bp mRNA linear EST 30-JAN-2003  
DEFINITION VBB603020215.R1 CSEQFXN41 testes Bos taurus cDNA, mRNA sequence.  
ACCESSION CB169246  
VERSION CB169246.1 GI:28155372  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 814)  
AUTHORS Adelson,D.L. and Gill,C.A.  
TITLE Bovine ESTs (Adelson and Gill)  
JOURNAL Unpublished (2003)  
COMMENT Contact: David L. Adelson  
Animal Breeding and Genetics  
Texas A&M University  
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471, USA  
Tel: 9798452616  
Fax: 9798456970  
Email: david.adelson@tamu.edu.  
FEATURES Location/Qualifiers  
source 1..814  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/tissue\_type="testes"  
/clone\_lib="CSEQFXN41 testes"  
/note="Organ: testes; Vector: pBluescript SK+; Site\_1: NotI; Site\_2: EcoRI; sequence 5' of the insert (5'-NNN...NNNinsert)  
the inserts (AAGAATTTCGATATCAAGCTTATCGATACCGTCGACCTCGAG. normalized Rd 1 library, sequenced 3' with M13R primer."  
ORIGIN  
Query Match 21.3%; Score 658.6; DB 14; Length 814;  
Best Local Similarity 90.3%; Pred. No. 1e-112;  
Matches 758; Conservative 0; Mismatches 54; Indels 27; Gaps 4;  
QY 1348 ATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATATTCATTCATCGATCGCATCCAG 1407  
Db 814 ATGGTTCAAGCAGCAGATCTTCTTCTGCCATTTCATATTCATTCATCGATCGCATCCAG 755  
QY 1408 GCCCAGAATGATACT-ACAAAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTTGT 1466  
Db 754 GCCCAGAATGACACTAACAAAAGGAGACCATCCCATTCATGATGGGTTTTGAGCCACTTGT 695  
QY 1467 GAACCAGAGGCTACTTCCACCTTCCCTCGATATGCAAAAAATAATTAAAGGGAGA 1526

Db 694 TAACCAAGACTACTTCCACCCACCTTCCCTCGATATGCAAAAATAATAAAAGAGAGA 635

QY 1527 AATGGTGAACACTATTTTGCAGATTAAATAGATAGATAAATAAAACTGTCTGTGAGGTTGTCAA 1586

Db 634 AATGGTCAACTATTTTGCAGATTGATAGATAGATAAATAAAACTGTCTGTGAGGTTGTCAA 575

QY 1587 TTAAACAAATTACATTGTATCCTGGATTTTCTCTGTAATTTAGTGAACAGTCACCATG 1646

Db 574 TTAAACAAATTACATTGTATCCTGGATTTTCTCTGTAATTTAGTGAACATCACCTTG 515

QY 1647 TGTCTTTCAAGATCTCTGTACAAACCACCTTTCCTGGTGGATAACAAAAGGTCCTTGG 1706

Db 514 TGTCTTTTCGAGATCTCTATTACAAACCACCTTTCCTGGTGGATAACAAAAGGTCCTTGG 455

QY 1707 AACTCATCTCATGCAAGACATGCTGAAAGATGCACCTTCGGTCTTTTGTTCAGATCCTCCGA 1766

Db 454 AATCCATCTCATGCAAGACATGCTGAAAGATGCTCTTCGGTCTTTTGTTCAGTCTCCGG- 396

QY 1767 GTGCTTCCCCCAAGTGCTACCTATATAATAATCACCCAGGCTAAGGACTGTATCGACTCC 1826

Db 395 -TGCTCTCCCCAAAGTGCTGCCCTATATAATAATCACCCAGGCTAAGGACTGTATTGACTCT 337

QY 1827 TTTGTFACCTCACTGTGTTTCGGCCATTCTGTAGTCTTATTTCAGATCCATGGACATAACAGG 1886

Db 336 TTTGTFACCTCACTGTGTTTCGGCCATTCTGTAGTCTTATTTCAGATCCATGGACATAACAGG 277

QY 1887 GCTCGACAGAGATAAGCTTGGTCATATTTCTTGAGGAATTTGCCACCTTGCAGGATGAG 1946

Db 276 GCTCGCAGAGAGATAAACTCGGTCATATTTCTTGAGGAATTTGTACCTTGCAGGATG-- 219

QY 1947 TTTATGACATTTTATTTTAATAGCAGAGAAGTTGTATGCAGCGCTTCACACCATGCTGT 2006

Db 218 -----AGGCAGAGAAGTTGTATGCAGCGCTTCATACTATGCTGT 180

QY 2007 TGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGTACCTGGGTCTCTTTACCAATA 2066

Db 179 TGAACAGGAGCCCCACAGGCGCATTTGGCCTGTTTAGTACCTGGGTCTCTTTACCAATA 120

QY 2067 ACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTGAATTGGAACCTCTA-CAGTATG 2125

Db 119 ACCTTCGAATTATGATACAGTATCTCTTAAGTGGCTTCGAATTGGAGCTCTACCAGCATG 60

QY 2126 CACGAGTACTATTACATATATTGGTATCTCTCTGAATTCCTTTACGCATGGTTGATGTC 2184

Db 59 CACGAGTATTATTACATATATTGGTATCTCTCTGAATTTCTTTATGCTGGTTGATGTC 1

RESULT 6

AK050805

LOCUS AK050805 3071 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone: D030020M24 product: CORNEAL WOUND HEALING RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.

ACCESSION AK050805

VERSION AK050805.1 GI:26094130

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

11042159

3

REFERENCE

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

4

REFERENCE

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE

AUTHORS 6 (bases 1 to 3071)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physiscal and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/. Location/Qualifiers

FEATURES

source

1..3071

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:D030020M24"

/db\_xref="MGI:2418664"

/db\_xref="taxon:10090"

/clone="D030020M24"

/tissue\_type="whole body"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="9 days embryo"

1..3071

misc\_feature

/note="CORNEAL WOUND HEALING RELATED PROTEIN homolog [Rattus norvegicus] (SPTR|Q9J101, evidence: FASTY, 96.2%ID, 43.5%length, match=948)"

ORIGIN

Query Match 21.1%; Score 653.8; DB 11; Length 3071;



Best Local Similarity 88.1%; Pred. No. 7.2e-112; Matches 736; Conservative 0; Mismatches 92; Indels 7; Gaps 2;		
QY	1316	AGACCAGTCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTG 1375
Db	210	AGACCAGTCTGTTGCAGAGGCTCAGAAACTGATGGTCCAGGCGCAGACCTTCTTCTG 269
QY	1376	CCATTCAATAAFTCATTCGATCGCATGCGCATCCAGGCCCAGAAATGATATACAAAAGGAGATC 1435
Db	270	CCATTCCACACCTCATTTGCACCACGGCATCCAGGCTCAGAATGGCACTACCAAAGGAGACC 329
QY	1436	ATCCAATTATGATGGGTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTACCTTCC 1495
Db	330	ATCCAATTATGATGGGTTTGAAGCCCTTGTGAACCAGAGATTACTTCCACCCACCTTCC 389
QY	1496	CTCGATATGCAAAAATAATTAAAGGGBAAGAAATGGTGAACCTATTTGCAAGATTAATAG 1555
Db	390	CTCGCTATGCAAAAATAATTAAAGAGAGAGAAATGGTCAACTATTTCTCAAGATTAATAG 449
QY	1556	ATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTTGATCCTGGATT 1615
Db	450	ACAGATAAAAACTGTCTGTGAAGTCTGTGAACCTTACCAAATTACACTGTATCCTGGATT 509
QY	1616	TTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAACCA 1675
Db	510	TTTTCTGTGAATTTAGTGAGCAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAACCA 569
QY	1676	CTTTCCTGGTGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAG 1735
Db	570	CTTTCCTGGTGATAACAAAAAGGTCTTTGGCACTCATCTCATGCAAGACATGGTGAAG 629
QY	1736	ATGCACCTCGGTCTTTTGTGAGATCCTCCGAGTGTCTTCCCCCAAGTGCTACCTATATAA 1795
Db	630	ATGCTCTGGGTCTTCGTGAG--TCCTCCGGTGCTCTCCCCCAAGTGTGCCTATATAA 687
QY	1796	TAATCACAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTCCGCCATTCTG 1855
Db	688	TAATCACAGGCTAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTCCGCCGTTTTG 747
QY	1856	TAGTCTTATTCCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATAT 1915
Db	748	TAGTCTGTTCCAGATCCATGGCCATAACAGGGCTCGGCAGAGAGACAAGCTTGGTCACAT 807
QY	1916	TCITTAGGAATTTGCCACCTTGAGGATGAGTTTATGACATTTTATTTAATAGGCAGAG 1975
Db	808	TCITTAGGAGTTTGCTACCTTGCAAGATGAGT-----AACAAFTGTTTGCAGGCAGAG 862
QY	1976	AAGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTG 2035
Db	863	AAGTTGATGCAGCGCTTCATACTATGTTGTTGAAACAGGAGCCTCAGAGACAACATCTG 922
QY	2036	GCCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGCAATTATGATACAGTACCTTCTA 2095
Db	923	GCCTGTTTAGGAACCTGGGTTCTTTACCATAACCTCCGGATTATGATCCAGTATCTGCTC 982
QY	2096	AGTGGCTTTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTGGT 2150
Db	983	AGTGGCTTTGAGCTGGAGCTGTACAGCATGCATGAGTACTACTACATCTACTGGT 1037
RESULT 7		
BG547889		
LOCUS		
DEFINITION		
602576187F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4704212 5',		
mRNA sequence.		
ACCESSION		
BG547889		
VERSION		
BG547889.1 GI:13546554		
KEYWORDS		
EST.		
SOURCE		
Homo sapiens (human)		
ORGANISM		
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE		
1 (bases 1 to 768)		

AUTHORS TITLE JOURNAL COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LCM1543 row: n column: 21 High quality sequence stop: 738.	
	Location/Qualifiers	
	1. .768	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
FEATURES source	/db_xref="taxon:9606"	
	/clone="IMAGE:4704212"	
	/lab_host="DH10B (T1 phage-resistant)"	
	/clone_lib="NIH_MGC_77"	
	/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgcctcgcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."	
ORIGIN	Query Match 21.1%; Score 653.6; DB 12; Length 768;	
	Best Local Similarity 96.9%; Pred. No. 8.8e-112;	
	Matches 698; Conservative 0; Mismatches 19; Indels 3; Gaps 3;	
	QY 2149 GTATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTC 2208	
	Db 1 GTATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTC 60	
QY	2209	TCAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGTAGTAGTAAAAAAC 2268
Db	61	TCAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGTAGTAGTAAAAAAC 120
QY	2269	AAAGAAAAAAGAAAGTTCCGCCATTGAGCCGAGAGATCACAATGAGCCAAGCATATCA 2328
Db	121	AAAGAAAAAAGAAAGTTCCGCCATTGAGCCGAGAGATCACAATGAGCCTAGCATATCA 180
QY	2329	GAAACATGTGCTGGAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACG 2388
Db	181	GAAACATGTGCTGGAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACG 240
QY	2389	TAAACCGAAGTTTGAGCTTGATAGTGAACAAGTTTCGGTATGAACACAGGTTTGCTCCATT 2448
Db	241	TAAACCGAAGTTTGAGCTTGATAGCGAACAAGTTTCGGTATGAACACAGGTTTGCTCCATT 300
QY	2449	CAACAGTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCT 2508
Db	301	CAACAGTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCT 360
QY	2509	CAATAA-ATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACT 2567
Db	361	CAATAACATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGGCAGCTAGTAAGCACT 420
QY	2568	TTCAACAGGCAAAAATGATATTGGAAAATATTTCCTAACCCGGACCATGAGGTTAATAGAA 2627
Db	421	TTCAACAGGCAAAAATGATATTGGAAAATATTTCCTAACCCGGACCATGAGGTTAATAGAA 480
QY	2628	TTTTAAAGGTTGCCAAACCCCAACTTTTGTGGTTATGAAGTTATTGGCAGGAGACACAAAA 2687
Db	481	TTTTAAAGGTTGCCAAAGACCAACTTTTGTGGTTATGAAGTTATTGGCAGGAGACACAAAA 540

QY 2688 AGGAATCTAAAGTTCCTCCTGAATTGATTTCTCTGCTCATAAAATATTTTCCCTGTGTGA 2747  
|||||  
Db 541 AGGAATCTAAAGTTCCTCCTGAATTGATTTCTCTGCTCATAAAATATTTTCCCTGTGTGA 600  
|||||  
QY 2748 AACTTGTTTGAGAGAGACTGGGGAGGTGGCCAT-AAAGGGGCAGAGTCTTCTTTTCAGACC 2806  
|||||  
Db 601 AACCTTGTTGAGAGAGACTGGGGAGGTGGCCATCAACGGGGCAGAGTCTTCTATCAGACC 660  
|||||  
QY 2807 CAACTCTTAGAGGGGCACATCACCAGGCTCCACATCACGGGAAGTGAGATGGATTCTTGG 2866  
|||||  
Db 661 CAATTCTTAGAGGGCCATTACAGAGGCTCCA-ATCACGGGAAGTGAGATGGATTCTCTGGG 719  
|||||

RESULT 8  
BQ965085 957 bp mRNA linear EST 21-AUG-2002  
LOCUS  
DEFINITION AGENCOURT\_10052197 NIH MGC 134 Mus musculus cDNA clone  
IMAGE:6509503 5', mRNA sequence.  
ACCESSION BQ965085  
VERSION BQ965085.1 GI:22380563  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 957)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM14076 row: o column: 08  
High quality sequence stop: 632.  
Location/Qualifiers

FEATURES  
source  
1..957  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6509503"  
/tissue\_type="undifferentiated limb"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 134"  
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Cloned unidirectionally. Primer: Oligo dT. Average insert  
size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:  
this is a NIH\_MGC Library."

ORIGIN  
Query Match 20.9%; Score 648; DB 13; Length 957;  
Best Local Similarity 87.7%; Pred. No. 9.7e-111;  
Matches 719; Conservative 0; Mismatches 100; Indels 1; Gaps 1;  
QY 541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAG 600  
|||||  
Db 26 GCGCGGACGGCATCATGGTTATGAAAGCTACTGTAGACGACGATGCTTCGGGATGGGAG 85  
|||||  
QY 601 CTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCCAAGAT 660  
|||||  
Db 86 CTCGGGTCCCGGAAAAAATGGAAAAAGTAGCACAAAGCTGGGTGGACATAACCCAGGAT 145  
|||||  
QY 661 TTTGAGAAGCTTGTGCGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTTGGT 720  
|||||  
Db 146 TTTGAAGATGCTTGTGAGAGCTGAAGTTGGGAGAACTGCTTCATGATAAGCTGTTTGGT 205  
|||||  
QY 721 CTTTTGAAGCCATGTCTGTCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATT 780  
|||||

Db 206 CTTTTGAAGCCATGTCTGTCTATTGAAATGATGGATCCCTAAGATGGATCCCGGTATGATC 265  
|||||  
QY 781 GGAAACCAAGTTAATCGAAAAAGTTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 840  
|||||  
Db 266 GGGAACCAAGTGAAATAGAAAAAGTTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT 325  
|||||  
QY 841 AAAATTAAAGATCTCACCTTGCCCTGAACCTGATAGGGATTATGGATACATGTTTTTGTGTGT 900  
|||||  
Db 326 AAAATTAAAGACCTCAGCTGCCTGAACTGATAGGAATAATGGACACCTGTTTCTGTGTGT 385  
|||||  
QY 901 TTGATAACGTGGTTAGAAGCCATTTCACCTGGGCACAGACAGTATTTACGTGCCCTTTACATT 960  
|||||  
Db 386 TTGATCACATGGCTCGAAGGCCATTCTTGGGCACAGACAGTGTTTACGTGCCCTTTACATT 445  
|||||  
QY 961 CATAATCCAGACTTTTATAGAAGATCCTGCTATGAAGCCTTTTGTCTCTGGGAATCTTGAAA 1020  
|||||  
Db 446 CATAATCCCAGACTTCATAGAAGATCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTGAAG 505  
|||||  
QY 1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTT 1080  
|||||  
Db 506 ATCTGCGACATTGCAACGGAAAAAGTAAATAAAGCTGCTGTTTTTCGAAGAGGAAGATTTC 565  
|||||  
QY 1081 CAGTCAATGACTTATGGATTTAAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGC 1140  
|||||  
Db 566 CAGTCAATGACATACGGGATTTAAAAATGGCCAAACAGTGTGACAGATCTCCGAGTTACAGGG 625  
|||||  
QY 1141 ATCTAAAAAGATGTGGAGGATGACATGCAAAAGAAAGAGTAAAGAGTACTCTCGAAGTCGACAA 1200  
|||||  
Db 626 ATGCTAAAGGATGTGGAAGATGATCTGCAAGGCGAGTAAAGAGTACTCGAAGTCGACAA 685  
|||||  
QY 1201 GGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTTCAGC 1260  
|||||  
Db 686 GGAGAAGAGAGAGATCCCGAGGTGGAACCTAGAACACCAGCAGTGTCTGGCAGCATTCAGC 745  
|||||  
QY 1261 AGAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAGAGACC 1320  
|||||  
Db 746 AGAGTGAAGTTTCACGCGAGTGTGCTCACAGTGTCTCATAGGCTTTTACTAAG-AAGAGAAC 804  
|||||  
QY 1321 AGTGCTGTTGCAGAAAGCTCAAAAAATTGATGGTTTCAGCAG 1360  
|||||  
Db 805 AGTGCTGTTGCAGAAAGGCTCAAAAAACTGATGGGCCCGCG 844  
|||||

RESULT 9  
AI790514/c  
LOCUS AI790514 828 bp mRNA linear EST 02-JUL-1999  
DEFINITION ul02b03.x1 Sugano mouse kidney mkia Mus musculus cDNA clone  
IMAGE:2064845 3' similar to WP:T23B12.4 CE14032 ;, mRNA sequence.  
ACCESSION AI790514 GI:5338230  
VERSION AI790514.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 828)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Other\_ESTs: ul02b03.y1  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:994033  
Possible reversed clone: similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 509.  
Location/Qualifiers  
FEATURES  
source  
1..828  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2064845"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Sugano mouse kidney mkia"  
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."

ORIGIN  
Query Match 20.7%; Score 641.4; DB 9; Length 828;  
Best Local Similarity 86.4%; Pred. No. 1.7e-109;  
Matches 716; Conservative 0; Mismatches 112; Indels 1; Gaps 1;

QY 673 TGTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTGTGCTCTTTTGAAGCC 732  
Db 828 TGTCGAGAGCTGAAGTTGGAAGACTGCTTCATTGATAAGCTGTTTCTTTTGAAGCC 769

QY 733 ATGCTCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTTGGAACCAAGTT 792  
Db 768 ATGCTCTGCTATTGAAATGATGATTCNTAAGATGGATGCGGGTATGATCNGAACCAAGTG 709

QY 793 AATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAATTAAGAT 852  
Db 708 AATAGNAAAGTTCTCAATTTTGANCAAGCTATCAAGGATGGCAC-ATTAAATTTAAAGAC 650

QY 853 CTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTTTGTGCTTTTGATAACGTGG 912  
Db 649 CTCAGCCTGCNTGAACTGATAGGAATAATGGACACCTGTTTCTGCTGTTTGATCACAATGG 590

QY 913 TTAGAAGCCATTCACTGGCACAGACAGTATTTACGTGCCCTTTACATTCATAATCCAGAC 972  
Db 589 CTCGAAGCCATTCTTGGCACAGACAGTGTTTACGTGCCCTTTACATTCATAATCCCGAC 530

QY 973 TTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGCGAATCTTGAATACTGTGACATT 1032  
Db 529 TTCATAGAAGATCCTGCCATGAAAGCTTTTGTCTCTGGGAATCTTGAAGATCTGCGACATT 470

QY 1033 GCAAGGAAAAAGTAAATAAGCTGCTGTTTTTTGAAGAGGAAGATTTTCAGTCAATGACT 1092  
Db 469 GCACGGAAAAAGTAAATAAGCTGCTGTTTTTCGAAGAGGAAGATTTCCAGTCAATGACA 410

QY 1093 TATGGATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAAAGAT 1152  
Db 409 TACGGATTTAAATGGCCCAACAGTGTGACAGATCTCCGAGTTACAGGGATGCTAAAGGAT 350

QY 1153 GTGGAGGATGACATGCAAAAGAGAGTAAGAGTACTCGAAGTCGACAAGGAGAGAAGA 1212  
Db 349 GTGGAAGATGATCTGCAAAGCGCAGTAAAGAGTACTCGAAGTCGACAAGGAGAGAAGA 290

QY 1213 GATCCAGAAGTTGAAC TAGAACCAACAATGTTTAGCAGTATTCAGCAGAGTGAAATTT 1272  
Db 289 GATCCAGAGGTCGAAC TAGAACCCAGCAGTGCTTGGCAGCATTTCAGCAGAGTGAAATTC 230

QY 1273 ACTCGTGTGTACTACAGTGTCTTATAGCCTTTACTAAGAAAGAGACCAGCTGCTGTGCA 1332  
Db 229 ACGCGAGTGTGCTACAGTGTCTCATAGCCTTTACTAAGAAAGAGACCAGTGTGTGCA 170

QY 1333 GAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCAATAATTCATTG 1392  
Db 169 GAGGCTCAGAAAATTGATGGTCCAGGCGCAGACCTTCTTTCTGCCATTACACCTCATTG 110

QY 1393 CATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGATCATCAATTATGATGGGT 1452  
Db 109 CACCACGGCATCAAGGCTCCGAATGGCACTACCAAAGGAGACCATCCAATTAAAATGGGT 50

QY 1453 TTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTACCTTCCCCTCGAT 1501  
Db 49 TTTGAGCCCTTGTTAACCAGAGATTACTTCCCCCGCTTCCCCTCGNT 1

RESULT 10  
CD808079  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 786)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1..786  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30540139"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 15.5,16.5,17.5,18.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_GW0"  
/note="Organ: Eye; Vector: pYX- Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CTGCGTCCCTC. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

FEATURES  
source  
ORIGIN



Query Match		20.6%;	Score 636.8;	DB 14;	Length 786;
Best Local Similarity		88.7%;	Pred. No. 1.2e-108;		
Matches 698;		Conservative 0;	Mismatches 87;	Indels 2;	Gaps 1;
QY	1150	GATGTGGAGGATGACATGCAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGAAGAA	1209		
Db	2	GATGTGGAAGATGATCTGCAAAGGCGAGTAAAGAGTACTCGAAGTCGACAAGGAGAAGAG	61		
QY	1210	AGAGATCCAGAAAGTTGAAC TAGAACACCAACAATGTTTAGCAGTATTTCAGCAGAGTGAAA	1269		
Db	62	AGAGATCCAGAGGTCGAAC TAGAACACCAACAGCAGTGCTTGGCAGCATTCAGCAGAGTGAAG	121		
QY	1270	TTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAAAGAGACCAGTGTGTT	1329		
Db	122	TTACGCGGAGTGTGCTCACAGTGTCTCATAGCCTTTACTAAGAAAGAGACCAGTGTGTT	181		
QY	1330	GCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCAATAATTCA	1389		
Db	182	GCAGAGGCTCAGAAACTGATGGTCCAGGCGGCAGACCTTCTTTCTGCCATTTCACACCTCA	241		
QY	1390	TTGCATCATGGCATCCAGGCCAGAGATGATACTACAAAAGGAGATCATCCAATTATGATG	1449		
Db	242	TTGCACCACGGCATCCAGGCTCAGATGGCACTACCAAAGGAGACCATCCAATTATGATG	301		
QY	1450	GGTTTGAACCCCTTGTGAACACGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAA	1509		
Db	302	GGTTTGAAGCCCTTGTGTTAACGACAGAGATTACTTCCACCCACCTTCCCTCGTATGCAAAA	361		
QY	1510	ATAATTAAAGGGAAGAAATGGTGAAC TATTTTGGCAAGATTAATAGATAGATAATAAAACT	1569		
Db	362	ATAATTAAAGAGAGAATAATGGTCAACTATTTCTCAAGATTAATAGACAGATAATAAAACT	421		
QY	1570	GTCGTGAGGTTGTGAATTTAACAATTTACATTTGTATCTCGGATTTTTCTGTGAATTT	1629		
Db	422	GTCTGTAAGTCGTGAACCTTACAAAACCTTACACTGTATCCTGGATTTTTCTGTGAATTT	481		
QY	1630	AGTGAACAGTCACCATGTGTTCTTCAAGATCTCTGTTACAAACCACTTTCCTGGTGGAT	1689		
Db	482	AGTGAGCAGTCACCATGTGTTCTTCAAGATCTCTGTTACAAACCACTTTCCTCGTGGAT	541		
QY	1690	AACAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCGGTCT	1749		
Db	542	AACANAAGGTCTTTGGCACTCATCTCATGCAAGACATGGTGAAGATGCTCTCGGGTCC	601		
QY	1750	TTTGTGATCCTCCGAGTGCTTTCCCCCAAGTGCTACCTATATATAATCACCAGGCTA	1809		
Db	602	TTCTGTGAG--TCCTCCGGTGCTCTCNCCCAAGTGTGGCCTATATAATAATCACCAGGCTA	659		
QY	1810	AGGACTGTATCGACTCCTTTTGTTTACTCACTGTGTTTCGGCCATTTCTGTAGTCTTATTGAG	1869		
Db	660	AGGACTGTATCGACTCCTTTTGTTTACTCACTGTGTTTCNGCGGTTTGTAGTCTTTGTTT	719		
QY	1870	TCCATGGACATAAACAGGGCTCGACAGAGAGATAAGCTTGGTGCATATTTCTTGAGGAATTTG	1929		
Db	720	TCCATGGNNCATACAGGGCTCGGCAGAGAGACAAGCTTGGTCACATTTCTTGAGGAGTTTG	779		
QY	1930	CCACCTT 1936			
Db	780	CTACCTT 786			

RESULT 11	BM979511/c				
LOCUS	BM979511	616 bp	mRNA	linear	EST 21-FEB-2003
DEFINITION	UI-CF-DUI1-adr-h-18-0-UI.s1	UI-CF-DUI1	Homo sapiens	cdna clone	
ACCESSION	BM979511				
VERSION	BM979511.1	GI:19600035			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				

REFERENCE

1 (bases 1 to 616)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
COMMENT  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES

Location/Qualifiers  
1..616  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI1-adr-h-18-0-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-DUI1"  
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-CF-DUI1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Epithelial Cells The  
library was constructed according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT7T3-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dT)18 tail. The sequence tag for this  
library is GGCTGTAGGC.  
TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368  
TAG\_LIB=UI-CF-DUI1  
TAG\_SEQ=GGCTGTAGGC"

ORIGIN

Query Match		19.7%;	Score 610;	DB 12;	Length 616;
Best Local Similarity		100.0%;	Pred. No. 1.3e-103;		
Matches 610;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2487	AGTTCAAGGAAATGCTGTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGT	2546		
Db	616	AGTTCAAGGAAATGCTGTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTCCTGAACTGT	557		
QY	2547	ATGTGGCAGCTAGTAAGCATTTC AACAGGCAAAAAATGATAATGGAAAAATATTCCTAACC	2606		
Db	556	ATGTGGCAGCTAGTAAGCATTTC AACAGGCAAAAAATGATAATGGAAAAATATTCCTAACC	497		
QY	2607	CGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCAACTTTGTGGTTATGAAGT	2666		
Db	496	CGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCAACTTTGTGGTTATGAAGT	437		
QY	2667	TATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTC	2726		
Db	436	TATTGGCAGGAGGACACAAAAGGAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTC	377		
QY	2727	ATAAATATTTTCCTGTTGTGAAACTTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGG	2786		

Db 376 ATAAATATTTTCCTGTTGAAACTTGTGTTGAGAGAGACTGGGAGGTGGCCATAAAAGGG 317  
QY 2787 GCAGAGTCCTTCTTTCAGACCCAACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGG 2846  
Db 316 GCAGAGTCCTTCTTTCAGACCCAACTCTTAGAGGGCACATCACCAGGCTCCACATCACGGG 257  
QY 2847 AAGTGAGATGGATTCTTGGGTAACAACCTCATATAAGGAATACCTTTTAGTTGACAGCC 2906  
Db 256 AAGTGAGATGGATTCTTGGGTAACAACCTCATATAAGGAATACCTTTTAGTTGACAGCC 197  
QY 2907 TTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGTTTATTATTATGTTCCATGGAAGAA 2966  
Db 196 TTATATGACATGAATGAAAACTGCTGTTTTTAAAGTGTTTATTATTATGTTCCATGGAAGAA 137  
QY 2967 CTGGTCTTATTGAATGCATTGATGAACGTTATATGGTTTTTATTACAGATTAAATCACAAA 3026  
Db 136 CTGGTCTTATTGAATGCATTGATGAACGTTATATGTTTTTATTACAGATTAAATCACAAA 77  
QY 3027 TCATTTTTTATGAATGATTGAGTGAAAAATAGTGTTTATAAAGTTTAATAAATTTCTTGAC 3086  
Db 76 TCATTTTTTATGAATGATTGAGTGAAAAATAGTGTTTATAAAGTTTAATAAATTTCTTGAC 17  
QY 3087 AAAAAA AAAA 3096  
Db 16 AAAAAA AAAA 7

RESULT 12  
CF748928  
LOCUS  
DEFINITION  
UI-M-HJ0-cwi-1-06-0-UI.r1 NIH\_BMAP\_HJ0 Mus musculus cDNA clone  
IMAGE:30626477 5', mRNA sequence.  
CF748928  
CF748928.1 GI:37645273  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 721)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1. .721  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30626477"  
/tissue\_type="Upper Head"  
/dev\_stage="9.5 and 10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_HJ0"  
/note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated

FEATURES  
source

with EcoR I adaptor , digested with NotI and then cloned  
directionally into pYX-Asc vector . The library tag  
sequence located between the Not I site and the polyA tail  
is CGAACTGAAT. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."  
ORIGIN  
Query Match 19.5%; Score 602.6; DB 14; Length 721;  
Best Local Similarity 89.7%; Pred. No. 3e-102;  
Matches 647; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
QY 691 GGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAGCCATGTCCTGCTATTGAAATG 750  
Db 1 GGAGAACTGCTTCATGATAAGCTGTTGGTCTTTTGAAGCCATGTCCTGCTATTGAAATG 60  
QY 751 ATGGATCCCAAGATGGATGCTGGCATGATTGGAACCAAGTTAATCGAAAAAGTTCTCAAT 810  
Db 61 ATGGATCCTAAAGATGGATGCCGGTATGATCGGAACCAAGTGAATAGAAAAAGTTCTCAAT 120  
QY 811 TTTGAACAAGCTATCAAGGATGGCACTATTAAAAATTAAAGATCTCACCTTGCCTGAAC TG 870  
Db 121 TTTGAACAAGCTATCAAGGATGGCACCAITAAAAATTAAAGACCTCAGCCTGCCTGAAC TG 180  
QY 871 ATAGGATTATGGATACATGTTTTTGTCTGTTTGATAACCGTGGTTAGAAGGCCATTCACTG 930  
Db 181 ATAGGAATAATGGACACCTGTTTCTGCTGTTTGATCACATGGCTCGAAGGCCATTCTCTTG 240  
QY 931 GCACAGACAGTATTTACGTGCCCTTACATTCATAATCCAGACTTTATAGAAGATCCTGCT 990  
Db 241 GCACAGACAGTGTTTACGTGCCCTTACATTCATAATCCCGACTTCATAGAAGATCCTGCC 300  
QY 991 ATGAAGGCTTTTGCTCTGCGAATCTTTGAAAATCTGTGACATTCGAAGGGAAAAAGTAAAT 1050  
Db 301 ATGAAAAGCTTTTGCTCTGCGAATCTTTGAAGATCTGCGACATTCGACGGAAATGAAAT 360  
QY 1051 AAAGCTGCTGTTTTTTGAAGAGGAAGATTTTCAGTCAATGACTTATGGATTTAAAAATGGCT 1110  
Db 361 AAAGCTGCTGTTTTTCGAAGAGGAAGATTTCCAGTCAATGACATACGGATTTAAAAATGGCC 420  
QY 1111 AACAGTGTGACAGATCTTTCGAGTTACAGGCATGCTAAAAGATGTGTGAGGATGACATGCAA 1170  
Db 421 AACAGTGTGACAGATCTCCGAGTTACAGGGATGCTAAAGGATGTGGAAGATGATCTGCAA 480  
QY 1171 AGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGAAAGAGATCCAGAAAGTTGAACTA 1230  
Db 481 AGCCGAGTAAAGAGTACTCGAAGTCGACAAGGAGAGAGAGATCCAGAGGTCGAACTA 540  
QY 1231 GAACACCAACAATGTTTAGCAGTATTTCAGCAGAGTGAATTTACTCGTGTGTACTGACA 1290  
Db 541 GAACACCAGCAGTGTCTTGGCAGCATTCAGCAGAGTGAAGTTTACGCGAGTGTCTGCTCACA 600  
QY 1291 GTGCTTATAGCCTTTTACTAAAGAAAAGAGACCAGTGTGTTGCAGAAAGCTCAAAAATTGATG 1350  
Db 601 GTGCTCATAGCCTTTTACTAAGAAAGAGACCAGTGTGTTGCAGAGGCTCAGAAACTGATG 660  
QY 1351 GTTCAAGCAGCAGATCTTCTTTCTGCCATTTCATAATTTCATTCATGGCATCCAGGCC 1410  
Db 661 GTCCAGGGGGCAGACCTTCTTTCTGCCATTTCACCTCATTTGCAACCGGCATCCAGGCT 720  
QY 1411 C 1411  
Db 721 C 721

RESULT 13  
CA977646  
LOCUS  
DEFINITION  
AGENCOURT 11295472 NIH\_MGC\_164 Mus musculus cDNA clone  
IMAGE:30146905 5', mRNA sequence.  
CA977646  
CA977646.1 GI:27510300  
CA977646 900 bp mRNA linear EST 06-JAN-2003

KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 900)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. David Rowe and Dr. Mina  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDAM0063 row: f column: 02  
High quality sequence stop: 655.  
Location/Qualifiers  
1. .900  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30146905"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 164"  
/note="Vector: pCMV-SPORT6.1; Site\_1: EcoRV; Site\_2: NotI;  
Non-normalized full-length enriched library from pooled  
mouse embryonic limb, maxilla and mandible, day 10.5 and  
11.5 (size selected for the 0.5-1 kb fragments) Cloned  
directionally, priming method: Oligo-dT. cDNA enrichment:  
>1k bp, Average insert size 1.8k bp. Priming sequence:  
5'GACTAGTTCTAGATCGCGAGCGGCCGCC(T) 3'. Tissue contributed  
by, David Rowe. Library constructed by ResGen, Invitrogen  
Corp."  
ORIGIN  
Query Match 19.4%; Score 599.2; DB 14; Length 900;  
Best Local Similarity 85.5%; Pred. No. 1.3e-101;  
Matches 700; Conservative 0; Mismatches 94; Indels 25; Gaps 2;  
QY 1535 ACTATTTGCAAGATTAAATAGATAGATAAATAAACTGTCTGTGAGTTGTGAATTTAACAA 1594  
Db 3 ACTATTTCTCAAGATTAAATAGACAGATAAATAAACTGTCTGTGAAGTCGTGAACCTACCAA 62  
QY 1595 ATTTACATGTATCCTGGATTCTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTT 1654  
Db 63 ACTTACACTGTATCCTGGATTCTTCTGTGAATTTAGTGAAGTCACCATGTGTTCTTT 122  
QY 1655 CAAGATCTCTGTACAAACCACCTTTCTGGTGATACAAAAAGTCTTTTGGAACTCATC 1714  
Db 123 CAAGATCTCTGTACAAACCACCTTTCTGGTGATACAAAAAGTCTTTGGCACTCATC 182  
QY 1715 TCATGCAAGACATGGTGAAAGATGCACTTCGGTCTTTTGTGAGATCCTCCGAGTCTTTC 1774  
Db 183 TCATGCAAGACATGGTGAAAGATGCTCTCGGTCTCTTCGTGAG--TCCTCCGGTGTCTTC 240  
QY 1775 CCCCAGTGCTACCTATATAATAATACACAGGCTAAGGACTGTATCGACTCCTTTGTTAC 1834  
Db 241 CCCCAGTGTTGCCTATATAATAATACACAGGCTAAGGACTGTATCGACTCCTTTGTTAC 300  
QY 1835 TCACTGTGTTCCGCCATCTCTGTAGTCTTATTTCAGATCCATGGACATAACAGGCTCGACA 1894  
Db 301 TCACTGTGTTCCGCCGTTTGTAGTCTTGTTCAGATCCATGGCCATAACAGGCTCGGCA 360  
QY 1895 GAGAGATAAGCTTGTCATATTCTTGAGGAATTTGCCACCTTGCAGATGAGTTTATGAC 1954  
Db 361 GAGAGACAAGCTTGTCATATTCTTGAGGAGTTTGTACCTTGCAAGATG----- 410  
QY 1955 ATTTTATTTTATAGGCAGAGAAGGTTGATGAGCGGCTTCACACCATGCTGTGTAACAG 2014

Db 411 -----AGGCAGAGAAAGGTTGATGACGCGCTTCATACTATGTTGTAACAG 457  
QY 2015 GAAACCCCAAAGGCAACATTTGGCCTGTTTAGGTACCTGGGTCTTTACCATAACCTTCGC 2074  
Db 458 GAGCCTCAGAGACAACATCTGGCCTGCTTAGGAACCTGGGTTCTTTACCATAACCTCCGG 517  
QY 2075 ATTATGATACAGTACCTTCTAAAGTGGCTTTGAAATTGGAACCTCTACAGTATGCACGAGTAC 2134  
Db 518 ATTATGATCCAGTATCTGCTCAGTGGCTTTGAGCTGAGAGCTGTACAGCATGCATGAGTAC 577  
QY 2135 TATTACATATATTGGTATCTCTCTGAATTCCTTTTACGCATGGTTGATGTCAACATTCAGT 2194  
Db 578 TACTACATCTACTGGTACCTCTCCGAGTTCCTGTATGCATGGCTGATGTCAACCTGAGC 637  
QY 2195 CGTCCGATGGCTCTCAAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGCCGT 2254  
Db 638 CGCGCTGACGGCTCTCAGATGGCAGAGGAGGATCATGGANGAACAGCAGAAAGCCGC 697  
QY 2255 AGTAGTAAAAAAACAAAGAAAAAAGAAAGTTTCGCCATTGAGCCGAGAGATCACAATG 2314  
Db 698 AGCCGCCAAAAAACCAAGAAAAAAGAAAGTTTCGTCGTTGAGCCGAGAGATCACAATG 757  
QY 2315 AGCCAAGCATATCAGAACATGTGTGCTGGAATGTTTAAA 2353  
Db 758 AGCCAGGCCTATCAGAACATGTGTGCTGGGATGTTCAA 796  
RESULT 14  
BI602827 697 bp mRNA linear EST 07-SEP-2001  
LOCUS 603247187F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5294804 5',  
DEFINITION mRNA sequence.  
ACCESSION BI602827  
VERSION BI602827.1 GI:15495766  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 697)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11745 row: n column: 21  
High quality sequence stop: 643.  
Location/Qualifiers  
1. .697  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5294804"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/clone\_lib="NIH MGC\_96"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI  
(gtcgag); Oligo-dT primed using primer  
5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
insert size 2.3 kb and normalized to 5. This is a  
primary library enriched for full-length clones and  
constructed using the Cap-trapper method (Carninci, in  
preparation). Library constructed by M. Brownstein  
(NIMH/NHGRI, National Institutes of Health). Note: this is





Search completed: August 11, 2004, 04:18:40  
Job time : 8430.09 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:17 ; Search time 1241.45 Seconds  
(without alignments)  
10594.393 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttctcagaaactccagg.....atttttgacaaaaaaaaa 3096

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query			DB	ID	Description
	Score	Match	Length			
1	3096	100.0	3096	6	ABQ75303	Abq75303 Human lun
2	2977.2	96.2	3029	9	ADC30041	Adc30041 Human nov
C 3	2577.8	83.3	2668	5	ABV29823	Abv29823 Human pro
C 4	2577.8	83.3	2668	5	ABV23940	Abv23940 Human pro
5	2417.2	78.1	2488	6	AAL49929	Aal49929 Human mol
6	1799.2	58.1	2543	9	ADB53632	Adb53632 Primary r
7	717.4	23.2	719	6	ABS51309	Abs51309 cDNA enco
8	451.2	14.6	482	8	ACH33702	Ach33702 Human end
9	451.2	14.6	482	9	ADC31969	Adc31969 Human nov
10	350.8	11.3	398	7	ABX50281	Abx50281 Bovine ES
C 11	343.8	11.1	799	6	ABQ51659	Abq51659 Oligonucl
12	343.8	11.1	799	6	ABQ51658	Abq51658 Oligonucl
C 13	292.4	9.4	356	5	ABV15331	Abv15331 Human pro
C 14	288.8	9.3	439	5	ABV45182	Abv45182 Human pro
C 15	288.8	9.3	439	5	ABV36125	Abv36125 Human pro
16	274.2	8.9	799	6	ABQ51661	Abq51661 Oligonucl
C 17	274.2	8.9	799	6	ABQ51660	Abq51660 Oligonucl
C 18	230.2	7.4	303	5	ABV06162	Abv06162 Human pro
19	217	7.0	2515	4	ABL07143	AbL07143 Drosophil
C 20	140.4	4.5	151	6	ABQ75302	Abq75302 Human lun
C 21	90.6	2.9	3401	4	ABL05906	AbL05906 Drosophil
C 22	90.6	2.9	4847	4	ABL07142	AbL07142 Drosophil
C 23	87.8	2.8	815	6	ABQ75372	Abq75372 Human lun

24	66.6	2.2	2000	7	ADA71938	Ada71938 Rice gene
25	60	1.9	60	6	ABN43747	Abn43747 Human spl
C 26	58.6	1.9	600	6	ABQ52497	Abq52497 Oligonucl
27	58.6	1.9	600	6	ABQ52496	Abq52496 Oligonucl
28	58.2	1.9	3163	9	ADC87060	Adc87060 Human GPC
29	56.4	1.8	320	3	AAA38183	Aaa38183 Primer us
30	56.2	1.8	5452	9	ADC86736	Adc86736 Human GPC
31	56	1.8	451	6	ABT10409	Abt10409 Human bre
32	56	1.8	1117	9	ADC86688	Adc86688 Human GPC
33	56	1.8	1286	6	ABI99656	Abi99656 Mouse isc
C 34	55.4	1.8	365	5	ABV54856	Abv54856 Human pro
35	55.2	1.8	320	3	AAA38186	Aaa38186 Primer us
36	54.8	1.8	320	3	AAA38185	Aaa38185 Primer us
C 37	54.8	1.8	615	7	ACA23975	Aca23975 prokaryot
38	54.6	1.8	840	6	ABQ35494	Abq35494 Oligonucl
C 39	54.6	1.8	840	6	ABQ35495	Abq35495 Oligonucl
40	54.6	1.8	1416	7	ABZ20967	Abz20967 Animal te
41	54.6	1.8	12733	6	ABK98631	Abk98631 Vector pE
42	54.6	1.8	12733	8	ACD13882	Acd13882 L. lactis
43	54.6	1.8	12739	6	ABK98592	Abk98592 Vector pE
44	54.6	1.8	12739	8	ACD13843	Acd13843 Plasmid p
45	54.2	1.8	318	3	AAA38184	Aaa38184 Primer us

ALIGNMENTS

RESULT 1  
ABQ75303  
ID ABQ75303 standard; cDNA; 3096 BP.  
XX  
AC ABQ75303;  
XX

DT 05-NOV-2002 (first entry)

DE Human lung specific nucleic acid sequence SEQ ID NO:42.

KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma; gene; chromosome 9; ss.  
XX

OS Homo sapiens.

XX WO200264788-A2.

PN 22-AUG-2002.

XX

PF 20-NOV-2001; 2001WO-US045080.

XX 20-NOV-2000; 2000US-0252054P.

PR (DIAD-) DIADEXUS INC.

XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;

DR WPI; 2002-657601/70.

XX

PT New lung specific nucleic acid useful in gene therapy or as vaccines for

PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung

PT diseases, as well as for diagnosing, monitoring or staging these

PT diseases.

XX Claim 1; Page 168-169; 282pp; English.

XX

CC The present invention describes an isolated lung specific nucleic acid

CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid

CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;



CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient  
CC with lung cancer, particularly by inducing an immune response against the  
CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs  
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,  
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-  
CC cancerous disease states in lung  
XX

SQ Sequence 3096 BP; 914 A; 606 C; 749 G; 827 T; 0 U; 0 Other;

Query Match 100.0%; Score 3096; DB 6; Length 3096;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTCCTCACGAAACTCCAGGCGTGTATAGGAAACATAAATCCGTTGTCAGGCAGCA	60
Db	1	TTTCCTCACGAAACTCCAGGCGTGTATAGGAAACATAAATCCGTTGTCAGGCAGCA	60
QY	61	GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGTAACCACTAAGGCTCTACG	120
Db	61	GCACGCTGTTGCTCTCGGAGCTTGGCTGCTCGTTCGTAACCACTAAGGCTCTACG	120
QY	121	CAAACTCCACGGTTTCCTTCGCTTCGCTCACCTTTCTAAGAAATTTCCAGAGGCA	180
Db	121	CAAACTCCACGGTTTCCTTCGCTTCGCTCACCTTTCTAAGAAATTTCCAGAGGCA	180
QY	181	GGCAGACGGGGCGGCTCTGAGACTCCGGGCTCCGGCTCTTCGGGAAACCGCCACTA	240
Db	181	GGCAGACGGGGCGGCTCTGAGACTCCGGGCTCCGGCTCTTCGGGAAACCGCCACTA	240
QY	241	CCAGGACTCCGACAGAGGGTGAAGAAAGATAAATTCGCGTCTCGCGATCGTCTCTAATC	300
Db	241	CCAGGACTCCGACAGAGGGTGAAGAAAGATAAATTCGCGTCTCGCGATCGTCTCTAATC	300
QY	301	TCGCGAAGAGAGGCGCGCCATCGGCCGAAACGAGGCGGTGGCGAGGGGGTG	360
Db	301	TCGCGAAGAGAGGCGCGCCATCGGCCGAAACGAGGCGGTGGCGAGGGGGTG	360
QY	361	TGCGGGGAGCGGAAAGTCCCGGAGTAAGGAGAGGGCGGGTCCGCGTCCCGG	420
Db	361	TGCGGGGAGCGGAAAGTCCCGGAGTAAGGAGAGGGCGGGTCCGCGTCCCGG	420
QY	421	GCATACGATGCGTGCACGCTGCGGTGCGGCTGAGAGGGAGGGGGCGGCGG	480
Db	421	GCATACGATGCGTGCACGCTGCGGTGCGGCTGAGAGGGAGGGGGCGGCGG	480
QY	481	GCCGAGCGGCGTCTGTTATTTCCGTTGTCGGACAGTGCCTGCGGCGGGTGACCA	540
Db	481	GCCGAGCGGCGTCTGTTATTTCCGTTGTCGGACAGTGCCTGCGGCGGGTGACCA	540
QY	541	GGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGG	600
Db	541	GGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGG	600
QY	601	CTCAGTATGCCAGAAAAATGGAGAAAGCAATACAACTGGGTGGACATTACCCAAGAT	660
Db	601	CTCAGTATGCCAGAAAAATGGAGAAAGCAATACAACTGGGTGGACATTACCCAAGAT	660
QY	661	TTTGAAGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAGCTATTGGT	720
Db	661	TTTGAAGAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAGCTATTGGT	720
QY	721	CTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATT	780
Db	721	CTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATT	780
QY	781	GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT	840
Db	781	GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATT	840
QY	841	AAAATTAAGATCTCACCTTGCCCTGAACTGATAGGATTTATGGATACATGTTTTGCTGT	900
Db	841	AAAATTAAGATCTCACCTTGCCCTGAACTGATAGGATTTATGGATACATGTTTTGCTGT	900

QY	901	TTGATAACGTTGGTTAGAACCCATTCACTGGCACAGACAGATATTTACGTGCCCTTTACATT	960
Db	901	TTGATAACGTTGGTTAGAACCCATTCACTGGCACAGACAGATATTTACGTGCCCTTTACATT	960
QY	961	CATAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGCGAATCTTGAA	1020
Db	961	CATAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTGTCTGCGAATCTTGAA	1020
QY	1021	ATCTGTGACATTGCAAGGGAATAAATAAGCTGCTGTTTTTGAAGAGGAAGATTTT	1080
Db	1021	ATCTGTGACATTGCAAGGGAATAAATAAGCTGCTGTTTTTGAAGAGGAAGATTTT	1080
QY	1081	CAGTCAATGACTTATGGAATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC	1140
Db	1081	CAGTCAATGACTTATGGAATTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC	1140
QY	1141	ATGCTAAAAGATGTGGAGTACATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAA	1200
Db	1141	ATGCTAAAAGATGTGGAGTACATGCAAGAAAGAGTAAAGAGTACTCGAAGTCGACAA	1200
QY	1201	GGAGAAAGAGATCCAGAAATTGAACCTAGAACACCAACAATGTTTAGCAGTATTTCAGC	1260
Db	1201	GGAGAAAGAGATCCAGAAATTGAACCTAGAACACCAACAATGTTTAGCAGTATTTCAGC	1260
QY	1261	AGAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCTTTTACTAAGAAAGAGAC	1320
Db	1261	AGAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCTTTTACTAAGAAAGAGAC	1320
QY	1321	AGTGTGTTGCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT	1380
Db	1321	AGTGTGTTGCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT	1380
QY	1381	CATAATTTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC	1440
Db	1381	CATAATTTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC	1440
QY	1441	ATTATGATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTTCCCTCGA	1500
Db	1441	ATTATGATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTTCCCTCGA	1500
QY	1501	TATGCAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGAAGATTAATAGATAGA	1560
Db	1501	TATGCAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGAAGATTAATAGATAGA	1560
QY	1561	ATAAAACTGTCTGTGAGTTGTGAATTTAACAATTTAATTGATGTTCTGTTCTGTTCTT	1620
Db	1561	ATAAAACTGTCTGTGAGTTGTGAATTTAACAATTTAATTGATGTTCTGTTCTGTTCTT	1620
QY	1621	TGTGAATTTAGTGAACAGTCACCATGTTCTTTCAAGATCTCTGTTTACAAACCACTTC	1680
Db	1621	TGTGAATTTAGTGAACAGTCACCATGTTCTTTCAAGATCTCTGTTTACAAACCACTTC	1680
QY	1681	CTGTGGATAACAAAAGGCTCTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA	1740
Db	1681	CTGTGGATAACAAAAGGCTCTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA	1740
QY	1741	CTTCGGTCTTTTGTGAGTCTTCCGAGTCTTCCCGGCTTCCCTATATAATAATC	1800
Db	1741	CTTCGGTCTTTTGTGAGTCTTCCGAGTCTTCCCGGCTTCCCTATATAATAATC	1800
QY	1801	ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCAGTGTGTTGCGGCTTCTGTAGTC	1860
Db	1801	ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCAGTGTGTTGCGGCTTCTGTAGTC	1860
QY	1861	TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGATTAAGCTTGGTCATATTCTTG	1920
Db	1861	TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGATTAAGCTTGGTCATATTCTTG	1920
QY	1921	AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTAAATAGGCAGAGAAGGT	1980
Db	1921	AGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTAAATAGGCAGAGAAGGT	1980

QY 1981 TGATGCAGCGCTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTG 2040  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1981 TGATGCAGCGCTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTG 2040  
QY 2041 TTTAGGTACCTGGTCTTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGG 2100  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2041 TTTAGGTACCTGGTCTTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGG 2100  
QY 2101 CTTTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTTGGTATCTCTCTGA 2160  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2101 CTTTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTTGGTATCTCTCTGA 2160  
QY 2161 ATTCCCTTTACGATGGTTGATGTCAACATTGAGTCGTCGCGATGGCTCAAATGGCAGA 2220  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2161 ATTCCCTTTACGATGGTTGATGTCAACATTGAGTCGTCGCGATGGCTCAAATGGCAGA 2220  
QY 2221 GGAAGGATATGGAAGAGCAGCAGAGAAAGCGCGTAGTAAAAAACAAGAAAAA 2280  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2221 GGAAGGATATGGAAGAGCAGCAGAGAAAGCGCGTAGTAAAAAACAAGAAAAA 2280  
QY 2281 GAAAGTTGCGCCATTGAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACATGTGTGC 2340  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2281 GAAAGTTGCGCCATTGAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACATGTGTGC 2340  
QY 2341 TGGAAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTT 2400  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2341 TGGAAATGTTAAAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTT 2400  
QY 2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGAT 2460  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGAT 2460  
QY 2461 GAACCCCGCGCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAATATAG 2520  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2461 GAACCCCGCGCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAATATAG 2520  
QY 2521 CCCTCCTCCTCAGTCTCCTGAAGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2521 CCCTCCTCCTCAGTCTCCTGAAGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
QY 2581 AATGATATTGGAATAATTTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2581 AATGATATTGGAATAATTTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGC 2640  
QY 2641 CAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2641 CAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
QY 2701 TCCTCCTGAATTTGATTTCTGCTCATAAATATTTTCTGTTGTGAAACTTGTGTTGAGA 2760  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2701 TCCTCCTGAATTTGATTTCTGCTCATAAATATTTTCTGTTGTGAAACTTGTGTTGAGA 2760  
QY 2761 GAGACTGGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTCAGACCCCAACTCTTAGAGGG 2820  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2761 GAGACTGGGGAGGTGGCCATAAAGGGCAGAGTCTTCTTTCAGACCCCAACTCTTAGAGGG 2820  
QY 2821 CACATCACCAGGCTCCACATCACGGAAGTGAGATGGATTTCTTGGGTAACTCATTA 2880  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2821 CACATCACCAGGCTCCACATCACGGAAGTGAGATGGATTTCTTGGGTAACTCATTA 2880  
QY 2881 TAAGGAATACTTTTAGTTTGACAGCTTATATGACATGAATGAAAACTGCTGTTTTAAG 2940  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2881 TAAGGAATACTTTTAGTTTGACAGCTTATATGACATGAATGAAAACTGCTGTTTTAAG 2940  
QY 2941 TGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGATGATGAACGTTATAT 3000  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2941 TGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGATGATGAACGTTATAT 3000  
QY 3001 GGTTTATTACAGATTTAATCACAATCATTTTTTATGAATGATGAGTGAAATAGTGT 3060  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
3001 GGTTTATTACAGATTTAATCACAATCATTTTTTATGAATGATGAGTGAAATAGTGT 3060  
QY 3061 TTATAAGGTTAATAAATTTCTTGACAAAAA 3096

Db 3061 TTATAAAGGTTAATAAATTTCTTGACAAAAA 3096  
RESULT 2  
ADC30041  
ID ADC30041 standard; cDNA; 3029 BP.  
XX  
AC ADC30041;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human novel cDNA sequence, SEQ ID NO:123.  
XX  
KW Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 9; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029271-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 24-SEP-2002; 2002WO-US030474.  
XX  
PR 24-SEP-2001; 2001US-0324631P.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX  
DR WPI; 2003-371981/35.  
DR P-PSDB; ADC31012.  
XX  
PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
XX  
PS Claim 1; SEQ ID NO 123; 1185pp; English.  
XX  
CC The invention relates to 971 novel human cDNA sequences (ADC29919-  
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the  
CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition; kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC contig sequences corresponding to the cDNA sequences of the invention  
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
CC -ADC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or



CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human cDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 3029 BP; 895 A; 593 C; 739 G; 802 T; 0 U; 0 Other;

Query Match 96.2%; Score 2977.2; DB 9; Length 3029;  
Best Local Similarity 99.1%; Pred. No. 0;

Matches 3023; Conservative 0; Mismatches 3; Indels 24; Gaps 2;

Qy	48	TCAGGACGACGATGACGCTGTTGGCTCTCGGAGCTTGGCTCGTTCGTCGCAACC	107
Db	1	TCAGGACGACGATGACGCTGTTGGCTCTCGGAGCTTGGCTCGTTCGTCGCAACC	60
Qy	108	ACTAAGGTCTACGCAACCTCCACGGTTTCCTTCGGCTTCGCGTACCTTTCTAAGAA	167
Db	61	ACTAAGGTCTACGCAACCTCCACGGTTTCCTTCGGCTTCGCGTACCTTTCTAAGAA	120
Qy	168	TTCCAGAGGGCAGCGCAGACGGGCGGGCTCTGAGACTCCGGCTCCGGCTCTTCCGG	227
Db	121	TTCCAGAGGGCAGCGCAGACGGGCGGGCTCTGAGACTCCGGCTCCGGCTCTTCCGG	180
Qy	228	GAACCGCCACTACCCAGGACTCCGACAGAGGGTGAAAAAGATAAACTTCGGTCTCGG	287
Db	181	GAACCGCCACTACCCAGGACTCCGACAGAGGGTGAAAAAGATAAACTTCGGTCTCGG	240
Qy	288	ATCGTCTCTAATCTCGCGAGAAGAGAGCGCGCCCATCGGCCGACGAGCGGTGGC	347
Db	241	ATCGTCTCTAATCTCGCGAGAAGAGAGCGCGCCCATCGGCCGACGAGCGGTGGC	300
Qy	348	GAGGAGGGGTGTGGCCGGGGAGCGCGAAGTCCCGGGAGTAAGGAGAGGGCGGGG	407
Db	301	GAGGAGGGGTGTGGCCGGGGAGCGCGAAGTCCCGGGAGTAAGGAGAGGGCGGGG	360
Qy	408	TCGCGGTCCCGGGCATAACGATCGGTGACCGTCCCGTCCGGTGGGCTGAGAGGGA	467
Db	361	TCGCGGTCCCGGGCATAACGATCGGTGACCGTCCCGTCCGGTGGGCTGAGAGGGA	420
Qy	468	GGGGCGGGCGGGCGGGCGGTTCGTTATTTCCGTGGTCCGGACAGTGCCTGGCGG	527
Db	421	GGGGCGGGCGGGCGGGCGGTTCGTTATTTCCGTGGTCCGGACAGTGCCTGGCGG	480
Qy	528	GCGGTGACACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCGTAGATGATACGA	587
Db	481	GCGGTGACACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCGTAGATGATACGA	540
Qy	588	TTCCAGGATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGA	647
Db	541	TTCCAGGATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGA	600
Qy	648	CATTACCAAGATTTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGA	707
Db	601	CATTACCAAGATTTTGAAGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGA	660
Qy	708	TAAGCTATTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGA	767
Db	661	TAAGCTATTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGA	720
Qy	768	TGCTGGCATGATGGAAACCAAGTTATCGAAAAAGTTCTCAATTTGAACAAGCTATCAA	827
Db	721	TGCTGGCATGATGGAAACCAAGTTATCGAAAAAGTTCTCAATTTGAACAAGCTATCAA	780
Qy	828	GGATGGCACTATTAAAAATTAAGATCTCACCTTGCCGAACTGATAGGATATGGATAC	887
Db	781	GGATGGCACTATTAAAAATTAAGATCTCACCTTGCCGAACTGATAGGATATGGATAC	840
Qy	888	ATGTTTTTGTGTTGATAACGTGGTTAGAGGCCATTACCTGGCACACAGTATTTTAC	947

Db	841	ATGTTTTTGTGTTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACGATATTTAC	900
Qy	948	GTGCCTTTTACATTAATCCAGACTTTTATAGAAGATCCCTGCTATGAAGGCTTTTGTCT	1007
Db	901	GTGCCTTTTACATTAATCCAGACTTTTATAGAAGATCCCTGCTATGAAGGCTTTTGTCT	960
Qy	1008	GGGAATCTTGAAAAATCTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTGA	1067
Db	961	GGGAATCTTGAAAAATCTGTGACATTCGAAGGGAAAAAGTAAATAAAGCTGCTGTTTTGA	1020
Qy	1068	AGAGGAAGATTTTCAGTCAATGACTTATGGATTTTAAATGGCTAACAGTGACAGATCT	1127
Db	1021	AGAGGAAGATTTTCAGTCAATGACTTATGGATTTTAAATGGCTAACAGTGACAGATCT	1080
Qy	1128	TCGAGTTACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAGAAGAGATAAGAGTAC	1187
Db	1081	TCGAGTTACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAGAAGAGATAAGAGTAC	1140
Qy	1188	TCGAAGTCGACAAAGGAGAAAGAGAGATCCAGAAAGTTGAAGTAACTAGAACCAACAATGTT	1247
Db	1141	TCGAAGTCGACAAAGGAGAAAGAGAGATCCAGAAAGTTGAAGTAACTAGAACCAACAATGTT	1200
Qy	1248	AGCACTATTTCAGCAGAGTGAATTTTACTCGTGTGTACTGACAGTGTCTATAGCCTTTAC	1307
Db	1201	AGCACTATTTCAGCAGAGTGAATTTTACTCGTGTGTACTGACAGTGTCTATAGCCTTTAC	1260
Qy	1308	TAAGAAAGAGACCAGTCTGTTCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCT	1367
Db	1261	TAAGAAAGAGACCAGTCTGTTCAGAAAGCTCAAAAATTTGATGGTTCAAGCAGCAGATCT	1320
Qy	1368	TCCTTCTGCCATTTCATAATTTCATTCATGATGGCATCCAGGCCAGAGATGATACAAA	1427
Db	1321	TCCTTCTGCCATTTCATAATTTCATTCATGATGGCATCCAGGCCAGAGATGATACAAA	1380
Qy	1428	AGGAGATCATCCAATTATGATGGTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACC	1487
Db	1381	AGGAGATCATCCAATTATGATGGTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACC	1440
Qy	1488	TACCTTCCCTCGATATGCAAAAATAATTAAGAGGGAAGAAATGGTGAACCTATTTGCAAG	1547
Db	1441	TACCTTCCCTCGATATGCAAAAATAATTAAGAGGGAAGAAATGGTGAACCTATTTGCAAG	1500
Qy	1548	ATTAATAGATAGAAATAAAACCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTGTAT	1607
Db	1501	ATTAATAGATAGAAATAAAACCTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTGTAT	1560
Qy	1608	CCCTGATTTTCTGTGAATTTAGTGAAACAGTCACCATGTGTCTTTCAAGATCTCTGTT	1667
Db	1561	CCCTGATTTTCTGTGAATTTAGTGAAACAGTCACCATGTGTCTTTCAAGATCTCTGTT	1620
Qy	1668	ACAAACCACTTCTCGTGGTGAATAACAAAAGGTCTTTGGAACTCATCTCATGCAAGACAT	1727
Db	1621	ACAAACCACTTCTCGTGGTGAATAACAAAAGGTCTTTGGAACTCATCTCATGCAAGACAT	1680
Qy	1728	GGTGAAGATGCACTTCG-GTCTTTTGTGAGATCCTCCGAGTGTCTTTCCCAAGTGCTA	1786
Db	1681	GGTGAAGATGCACTTCGCGTCTTTTGTGAGATCCTCCGAGTGTCTTTCCCAAGTGCTA	1740
Qy	1787	CCTATAATAATAATCACCAGGCTAAGGACTGATCGACTCCTTTGTTACTCACTGTGTTTCG	1846
Db	1741	CCTATAATAATAATCACCAGGCTAAGGACTGATCGACTCCTTTGTTACTCACTGTGTTTCG	1800
Qy	1847	GCCATTCTGTAGTCTTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCT	1906
Db	1801	GCCATTCTGTAGTCTTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCT	1860
Qy	1907	TGGTCATATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAA	1966
Db	1861	TGGTCATATTCTTGAGGAATTTGCCACCTTGCAGGATG-----	1898
Qy	1967	TAGGCAGAGAAGTTGATGCAGCGCTTCACACCATGCTGTTTGAACAGGAACCCCAAGG	2026
Db	1899	-AGGCAGAGAAGTTGATGCAGCGCTTCACACCATGCTGTTTGAACAGGAACCCCAAGG	1957



QY 2027 CAACATTTGGCCTGTTTAGGTACCTGGGTCCCTTACCATAAACCTTCGCATTATGATACAG 2086  
Db 1958 CAACATTTGGCCTGTTAGGTACCTGGGTCCCTTACCATAAACCTTCGCATTATGATACAG 2017  
QY 2087 TACCTTCTAAGTGGCTTTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATAT 2146  
Db 2018 TACCTTCTAAGTGGCTTTGAATTGGAACCTCTACAGTATGCACGAGTACTATTACATATAT 2077  
QY 2147 TGGTATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGC 2206  
Db 2078 TGGTATCTCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGC 2137  
QY 2207 TCTCAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAA 2266  
Db 2138 TCTCAAATGGCAGAGGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAA 2197  
QY 2267 ACAAGAAAAAAGAAAGTTCCGCCCATTTGAGCCGAGAGATCACAATGAGCCAAAGCATAT 2326  
Db 2198 ACAAGAAAAAAGAAAGTTCCGCCCATTTGAGCCGAGAGATCACAATGAGCCAAAGCATAT 2257  
QY 2327 CAGAACATGTGTGCTGGAATGTTTAAAAACCATGGTAGCAATTTGACATGGACGGCAAAGTA 2386  
Db 2258 CAGAACATGTGTGCTGGAATGTTTAAAAACCATGGTAGCAATTTGACATGGACGGCAAAGTA 2317  
QY 2387 CGTAAACCGAAGTTTGAGCTTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCA 2446  
Db 2318 CGTAAACCGAAGTTTGAGCTTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCA 2377  
QY 2447 TTCAACAGTGTGATGACCCCGCCGAGTGCACACTACTTACAGTTCAAGGAAATGTCAGAC 2506  
Db 2378 TTCAACAGTGTGATGACCCCGCCGAGTGCACACTACTTACAGTTCAAGGAAATGTCAGAC 2437  
QY 2507 CTCAATAAATATAGCCCTCCTCTCAGTCTCCTGAACTGATGTGGCAGCTAGTAAGCAC 2566  
Db 2438 CTCAATAAATATAGCCCTCCTCTCAGTCTCCTGAACTGATGTGGCAGCTAGTAAGCAC 2497  
QY 2567 TTTCAACAGGCAAAAATGATTTGGAAAAATATTCCTAACCCGACCATGAGGTTAATAGA 2626  
Db 2498 TTTCAACAGGCAAAAATGATTTGGAAAAATATTCCTAACCCGACCATGAGGTTAATAGA 2557  
QY 2627 ATTTTAAAGTTGCCAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGCACAAA 2686  
Db 2558 ATTTTAAAGTTGCCAAACCCAACTTTGTGTTATGAAGTTATTGGCAGGAGCACAAA 2617  
QY 2687 AAGGAATCTAAAGTTCTCCTGATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTG 2746  
Db 2618 AAGGAATCTAAAGTTCTCCTGATTTGATTTCTCTGCTCATAAATATTTTCTGTTGTG 2677  
QY 2747 AAACITTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGGGAGAGTCTTCTTTTCAGACC 2806  
Db 2678 AAACITTTTGAGAGAGACTGGGAAGGTGGCCATAAAGGGGAGAGTCTTCTTTTCAGACC 2737  
QY 2807 CAACCTTAGAGGGCACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGG 2866  
Db 2738 CAACCTTAGAGGGCACATCACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGG 2797  
QY 2867 GTAACAACTCATTATAAGGAATACTTTATGTTTCCATGGAAGAACTGGTCTATTGAATGCATT 2926  
Db 2798 GTAACAACTCATTATAAGGAATACTTTATGTTTCCATGGAAGAACTGGTCTATTGAATGCATT 2857  
QY 2927 CTGCTGTTTAAAGTGGTTTATATGTTTCCATGGAAGAACTGGTCTATTGAATGCATT 2986  
Db 2858 CTGCTGTTTAAAGTGGTTTATATGTTTCCATGGAAGAACTGGTCTATTGAATGCATT 2917  
QY 2987 GATGAACGTATATGTTTTTATACAGATTTAATCACAATCAATTTTTTATGAATGATTG 3046  
Db 2918 GATGAACGTATATGTTTTTATACAGATTTAATCACAATCAATTTTTTATGAATGATTG 2977  
QY 3047 AGTGAAATAGTGTTTATAAAGGTTAATAAATTTCTTGACAAAAA 3096  
Db 2978 AGTGAAATAGTGTTTATAAAGGTTAATAAATTTCTTGACAAAAA 3027

RESULT 3  
ABV29823/c  
ID ABV29823 standard; cDNA; 2668 BP.  
XX  
AC ABV29823;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 29814.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US005171.  
XX  
PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 6420; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 0 U; 6 Other;  
  
Query Match 83.3%; Score 2577.8; DB 5; Length 2668;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 2638; Conservative 0; Mismatches 7; Indels 26; Gaps 3;  
  
QY 426 CGCATGCGTGACCGCTCCCGTGGGCTGGGCTGAGAGGGAGGGGGCGGCCGA 485  
Db 2668 CGCATGCGTGACCGCTCCCGTGGGCTGGGCTGAGAGGGAGGGGGCGGCCGA 2609  
QY 486 GCGGCGTCGTTATTTCCGTGTCGGACAGTCGTCGGCGCGGGTGACCCACGGGAGA 545  
Db 2608 GCGGCGTCGTTATTTCCGTGTCGGACAGTCGTCGGCGCGGGTGACCCACGGGAGA 2549  
QY 546 AGTAGCATAAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAG 605  
Db 2548 AGTAGCATAAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAG 2489  
QY 606 TATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCAAGATTTTGA 665

Db ||||| 2488 TATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTTGA 2429

QY ||||| 666 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAAGCTATTTTGGTCTTTT 725

Db ||||| 2428 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTTCATGATAAAGCTATTTTGGTCTTTT 2369

QY ||||| 726 TGAAGCCCATGTCTGCTATTGAAATGATGGATCCCAAGATGAGTCTGGCATGATTGGAAA 785

Db ||||| 2368 TGAAGCCCATGTCTGCTATTGAAATGATGGATCCCAAGATGAGTCTGGCATGATTGGAAA 2309

QY ||||| 786 CCAAGTTAAATCGAAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGGCACATATTAAAT 845

Db ||||| 2308 CCAAGTTAAATCGAAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGGCACATATTAAAT 2249

QY ||||| 846 TAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTTTGGTCTGTTGAT 905

Db ||||| 2248 TAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTTTGGTCTGTTGAT 2189

QY ||||| 906 AACGTGGTTAGAAGGCCATTCACTGGCCACAGACAGTATTACGTGCCTTTACATTCATAA 965

Db ||||| 2188 AACGTGGTTAGAAGGCCATTCACTGGCCACAGACAGTATTACGTGCCTTTACATTCATAA 2129

QY ||||| 966 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCCTCGGGAATCTTGAAAACTG 1025

Db ||||| 2128 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCCTCGGGAATCTTGAAAACTG 2069

QY ||||| 1026 TGACATTCGAAGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAAGAGGAAGATTTTCAGTC 1085

Db ||||| 2068 TGACATTCGAAGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAAGAGGAAGATTTTCAGTC 2009

QY ||||| 1086 AATGACTTATGGATTTTAAATGGCTTAACAGTGTGACAGATCTTCGAGTTACAGGCAATGCT 1145

Db ||||| 2008 AATGACTTATGGATTTTAAATGGCTTAACAGTGTGACAGATCTTCGAGTTACAGGCAATGCT 1949

QY ||||| 1146 AAAAGATGTGGAGGATGACATGCAAGAAGAGATAAAGAGTACTCGAAGTCGACAAGGAGA 1205

Db ||||| 1948 AAAAGATGTGGAGGATGACATGCAAGAAGAGATAAAGAGTACTCGAAGTCGACAAGGAGA 1889

QY ||||| 1206 AGAAAGAGATCCAGAAAGTTGAACACTAGAACACCAACAAATGTTTAGCAGTATTTCAGCAGAGT 1265

Db ||||| 1888 AGAAAGAGATCCAGAAAGTTGAACACTAGAACACCAACAAATGTTTAGCAGTATTTCAGCAGAGT 1829

QY ||||| 1266 GAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCCTTTACTAAGAAAAGAGACCAATGTC 1325

Db ||||| 1828 GAAATTTACTCGTGTGTTACTGACAGTGTCTATAGCCCTTTACTAAGAAAAGAGACCAATGTC 1769

QY ||||| 1326 TGTGTCAGAAAGCTCAAAAATFGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1385

Db ||||| 1768 TGTGTCAGAAAGCTCAAAAATFGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1709

QY ||||| 1386 TTCATTGCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGATCATCCAAATTAT 1445

Db ||||| 1708 TTCATTGCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGGAGATCATCCAAATTAT 1649

QY ||||| 1446 GATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1505

Db ||||| 1648 GATGGGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC 1589

QY ||||| 1506 AAAAATAATTAAAAGGGAAGAAATGGTGAACTATTTTGCAAGATTAAATAGATAGATAAA 1565

Db ||||| 1588 AAAAATAATTAAAAGGGAAGAAATGGTGAACTATTTTGCAAGATTAAATAGATAGATAAA 1529

QY ||||| 1566 AACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATGTATCCTGGATTTTTCTGTGA 1625

Db ||||| 1528 AACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATGTATCCTGGATTTTTCTGTGA 1469

QY ||||| 1626 ATTTAGTGAACAGTCACCAATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGT 1685

Db ||||| 1468 ATTTAGTGAACAGTCACCAATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGT 1409

QY ||||| 1686 GGATAACAAAAAGGCTCTTTGGAACCTCATCTCATGCAAGACATGGTGAAAGATGCACTTCG 1745

Db ||||| 1408 GGATAACAAAAAGGCTTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACCTTCG 1349

QY ||||| 1746 GTCTTTTGTGAGATCCTCCGAGTGCCTTTCCCCCAAGTGTCTACCTATATAATAATCACAG 1805

Db ||||| 1348 GTCTTTTGTGAG--TCCTCCGGTGCCTTTCCCCCAAGTGTCTACCTATATAATAATCACAG 1291

QY ||||| 1806 GCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTGCGGCCATCTGTAGTCTTAFT 1865

Db ||||| 1290 GCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTGCGGCCATCTGTAGTCTTAFT 1231

QY ||||| 1866 CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA 1925

Db ||||| 1230 CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA 1171

QY ||||| 1926 TTTGCCACCTTGCAAGATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGTTTGATG 1985

Db ||||| 1170 TTTGCCACCTTGCAAGATG-----AGGCAGAGAAGTTTGATG 1134

QY ||||| 1986 CAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAACATTTTGGCCTGTTAG 2045

Db ||||| 1133 CAGCGCTTCACACCATGCTGTTGAAAACAGGAACCCCAAGGCAACATTTTGGCCTGTTAG 1074

QY ||||| 2046 GTACCTGGGTCTTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG 2105

Db ||||| 1073 GTACCTGGGTCTTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG 1014

QY ||||| 2106 AATTGGAACCTACAGTATGCACGAG-TACTATTACATATATTGGTATCTCTCTGAATTC 2164

Db ||||| 1013 AATTGGAACCTACAGTATGCACGAGATACTATTACATATATTGGTATCTCTCTGAATTC 954

QY ||||| 2165 CTTTACGCATGGTTGATGTCAAACATTGAGTCGTGCGATGGCTCTCAAATGGCAGAGGAA 2224

Db ||||| 953 CTTTACGCATGGTTGATGTCAAACATTGAGTCGTGCGATGGCTCTCAAATGGCAGAGGAA 894

QY ||||| 2225 AGGATAATGGAAGAGCAGAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAA 2284

Db ||||| 893 AGGATAATGGAAGAGCAGAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAA 834

QY ||||| 2285 GTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGA 2344

Db ||||| 833 GTTCGCCCATTTGAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTGCTGGA 774

QY ||||| 2345 ATGTTTAAAAACCATGGTAGCATTTGACATGGACCGCAAGTACGTAAACCGAAGTTTGAG 2404

Db ||||| 773 ATGTTTAAAAACCATGGTAGCATTTGACATGGACCGCAAGTACGTAAACCGAAGTTTGAG 714

QY ||||| 2405 CTTGATAGTGAAACAAAGTTCCGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGATGACC 2464

Db ||||| 713 CTTGATAGTGAAACAAAGTTCCGTATGAACACAGGTTTGTCTCCATTCAACAGTGTGATGACC 654

QY ||||| 2465 CCGCCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAAATAGCCCT 2524

Db ||||| 653 CCGCCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAAATAGCCCT 594

QY ||||| 2525 CCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAAATG 2584

Db ||||| 593 CCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAAGCACTTTCAACAGGCAAAAATG 534

QY ||||| 2585 ATATTGGAAAAATATTCCTAAACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAA 2644

Db ||||| 533 ATATTGGAAAAATATTCCTAAACCCGGACCATGAGGTTAATAGAAATTTTAAAGGTTGCCAAA 474

QY ||||| 2645 CCCAACTTTGTGGTTATGAAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAAGTTCCCT 2704

Db ||||| 473 CCCAACTTTGTGGTTATGAAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAAGTTCCCT 414

QY ||||| 2705 CCTGAATTTGATTTCTGTCTCATAAATATTTTCTGTGTGAAAATTTTGTGAGAGAGA 2764

Db ||||| 413 CCTGAATTTGATTTCTGTCTCATAAATATTTTCTGTGTGAAAATTTTGTGAGAGAGA 354

QY ||||| 2765 CTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTTCAGACCCCACTCTTAGAGGGGCACA 2824

Db ||||| 353 CTGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTTTCAGACCCCACTCTTAGAGGGGCACA 294



Qy 2825 TCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACACTCAATTATAAG 2884  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
293 TCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACACTCAATTATAAG 234  
Qy 2885 GAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTAAAGTGGT 2944  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
233 GAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGTTTAAAGTGGT 174  
Qy 2945 TTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTT 3004  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
173 TTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTT 114  
Qy 3005 TTATTACAGATTTAATCACAAATCATTTTTTTTATGAATGATGAGTGAAAAATAGTGTATT 3064  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
113 TTATTACAGATTTAATCACAAATCATTTTTTTTATGAATGATGAGTGAAAAATAGTGTATT 54  
Qy 3065 AAAGGTTAATAAATTTCTTGACAAAAAANA 3095  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
53 AAAGGTTAATAAATTTCTTGACAAAAAANA 23

RESULT 4  
ABV23940/c  
ID ABV23940 standard; cDNA; 2668 BP.  
XX  
AC ABV23940;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cdNA 23931.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynanamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US005171.  
XX  
PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX

PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 4423; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynanamic or pharmacogenomic marker  
XX  
SQ Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 0 U; 6 Other;  
Query Match 83.3%; Score 2577.8; DB 5; Length 2668;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 2638; Conservative 0; Mismatches 7; Indels 26; Gaps 3;  
Qy 426 CGCATGCGTGACGCTGCCGTGGGCTGGGCTGAGAGGGGCGGCGGCCGA 485  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2668 CGCATGCGTGACGCTGCCGTGGGCTGGGCTGAGAGGGGCGGCGGCCGA 2609  
Qy 486 GCGGCGTGGTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGGGGTGACCAACGGGAGA 545  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2608 GCGGCGTGGTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGGGGTGACCAACGGGAGA 2549  
Qy 546 AGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTGAGATGGGAGCTCAG 605  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2548 AGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTGAGATGGGAGCTCAG 2489  
Qy 606 TATGCCAGAAAAAATGGAGAAAAACAATACAACTGGGTGGACATTACCCAAGATTTGA 665  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2488 TATGCCAGAAAAAATGGAGAAAAACAATACAACTGGGTGGACATTACCCAAGATTTGA 2429  
Qy 666 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTATGATAAGCTATTTGGTCTTTT 725  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2428 AGAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTATGATAAGCTATTTGGTCTTTT 2369  
Qy 726 TGAAGCCATGTCGTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGAAA 785  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2368 TGAAGCCATGTCGTCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGATTGAAA 2309  
Qy 786 CCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGTATTAAT 845  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2308 CCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCAGTATTAAT 2249  
Qy 846 TAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTTTGTCTGTTGAT 905  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2248 TAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTTTGTCTGTTGAT 2189  
Qy 906 AACGTGGTTAGAAAGGCCATTCACTGGCACAGACAGATATTACGTGCCTTTACATTCATA 965  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2188 AACGTGGTTAGAAAGGCCATTCACTGGCACAGACAGATATTACGTGCCTTTACATTCATA 2129  
Qy 966 TCCAGACTTTATAGAAAGTCTGCTATGAAGGCTTTTGTCTGGGAATCTTGAAAACTG 1025  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2128 TCCAGACTTTATAGAAAGTCTGCTATGAAGGCTTTTGTCTGGGAATCTTGAAAACTG 2069  
Qy 1026 TGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC 1085  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2068 TGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC 2009  
Qy 1086 AATGACTTATGGATTTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT 1145  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
2008 AATGACTTATGGATTTTAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT 1949  
Qy 1146 AAAAGATGTGGAGGATGACATGCAAAAGAGATTAAGAGTACTCGAAGTCGACAGGAGA 1205  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1948 AAAAGATGTGGAGGATGACATGCAAAAGAGATTAAGAGTACTCGAAGTCGACAGGAGA 1889  
Qy 1206 AGAAAGAGATCCAGAAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAT 1265  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1888 AGAAAGAGATCCAGAAAGTTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAT 1829  
Qy 1266 GAAATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTACTAAGAAAAGAGACAGTGC 1325  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1828 GAAATTTACTCGTGTGTACTGACAGTGTCTTATAGCCTTTACTAAGAAAAGAGACAGTGC 1769  
Qy 1326 TGTTCAGAAAGCTCAAAAATGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1385  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1768 TGTTCAGAAAGCTCAAAAATGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATAA 1709  
Qy 1386 TTCATTGCATCATGGCATCCAGGCCAGAAATGATACTACAAAAGGAGATCATCCAATTAT 1445





XX (INCY-) INCYTE GENOMICS INC.  
PA Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;  
XX Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;  
PI Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;  
PI Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;  
PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;  
XX WPI; 2002-713453/77.  
DR P-PSDB; AAO19400.  
DR  
XX New human molecules for disease detection and treatment (MDDT), useful  
PT for diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant MDDT expression, e.g. cancer, AIDS, asthma, diabetes,  
PT hepatitis.  
XX  
PS Claim 11; Page 159-160; 177pp; English.  
XX  
CC The present invention relates to human proteins and coding sequences of  
CC molecules for disease detection and treatment MDDT. The sequences can be  
CC used in the treatment of diseases associated with the decreased  
CC expression or overexpression of MDDT, such as cell proliferative (cancer,  
CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,  
CC allergies, Addison's disease, asthma), developmental (dwarfism, renal  
CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,  
CC epilepsy) and cardiovascular (congestive heart failure, myocardial  
CC infarction, angina pectoris) disorders. The present sequence is a coding  
CC sequence of the invention  
XX  
SQ Sequence 2488 BP; 768 A; 461 C; 567 G; 692 T; 0 U; 0 Other;  
  
Query Match 78.1%; Score 2417.2; DB 6; Length 2488;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;  
  
QY 474 GGCGGCGCGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGCGGT 533  
Db 1 GGCGGCGCGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGTGGCGCGCGGT 60  
  
QY 534 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593  
Db 61 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120  
  
QY 594 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 653  
Db 121 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 180  
  
QY 654 CCAAGATTTGAAGAAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 713  
Db 181 CCAAGATTTGAAGAAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 240  
  
QY 714 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 773  
Db 241 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 300  
  
QY 774 CATGATTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833  
Db 301 CATGATTGGAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360  
  
QY 834 CACTATTAAATTAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTT 893  
Db 361 CACTATTAAATTAAGATCTCACCTTGCTGAACTGATAGGGATTATGGATACATGTTT 420  
  
QY 894 TTGCTGTTTGAATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTACGTGCCT 953  
Db 421 TTGCTGTTTGAATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTACGTGCCT 480  
  
QY 954 TTACATTCATAATCCAGACTTTATAGAAAGATCCTGCTATGAAGGCTTTTCTCTGGGAAT 1013  
Db 481 TTACATTCATAATCCAGACTTTATAGAAAGATCCTGCTATGAAGGCTTTTCTCTGGGAAT 540  
  
QY 1014 CTTGAAAAATCTGTGACATTGCAAGGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGA 1073

Db 541 CTTGAAAAATCTGTGACATTGCAAGGGAAGAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGA 600  
  
QY 1074 AGATTTTCAGTCAATGACTTATGGATTTAAAAATGGTAAACAGTGTGACAGATCTTCGAGT 1133  
Db 601 AGATTTTCAGTCAATGACTTATGGATTTAAAAATGGTAAACAGTGTGACAGATCTTCGAGT 660  
  
QY 1134 TACAGGCATGCTAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAG 1193  
Db 661 TACAGGCATGCTAAAGATGTGGAGGATGACATGCAAAAGAGAGTAAAGAGTACTCGAAG 720  
  
QY 1194 TCGACAAGGAGAAAGAGATCCAGAAAGTTGAACACTAGAACACCAACAATGTTTAGCAGT 1253  
Db 721 TCGACAAGGAGAAAGAGATCCAGAAAGTTGAACACTAGAACACCAACAATGTTTAGCAGT 780  
  
QY 1254 ATTCAGCAGAGTGAATTTACTCTGTGTGTTACTGACAGTCTTATAGCCTTTACTAAGAA 1313  
Db 781 ATTCAGCAGAGTGAATTTACTCTGTGTGTTACTGACAGTCTTATAGCCTTTACTAAGAA 840  
  
QY 1314 AGAGACCAGTCTGTTGCAGAAAGCTCAAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTC 1373  
Db 841 AGAGACCAGTCTGTTGCAGAAAGCTCAAAAAATTTGATGGTTCAAGCAGCAGATCTTCTTTC 900  
  
QY 1374 TGCCATTCAATAATTCATTGCATCATGGCATCCAGGCCCAGAAATGATACATAAAGGAGA 1433  
Db 901 TGCCATTCAATAATTCATTGCATCATGGCATCCAGGCCCAGAAATGATACATAAAGGAGA 960  
  
QY 1434 TCATCCAAATTATGATGGGTTTTTGAAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTT 1493  
Db 961 TCATCCAAATTATGATGGGTTTTTGAAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTT 1020  
  
QY 1494 CCCTCGATATGCAAAAAATAATTAAGGGAAGAAATGGTGAACATAATTTTGAAGATTAAT 1553  
Db 1021 CCCTCGATATGCAAAAAATAATTAAGGGAAGAAATGGTGAACATAATTTTGAAGATTAAT 1080  
  
QY 1554 AGATAGAATAAAAACTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATCCTGGA 1613  
Db 1081 AGATAGAATAAAAACTGCTGTGAGGTTGTGAATTTAAACAAATTTACATTTGATCCTGGA 1140  
  
QY 1614 TTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1673  
Db 1141 TTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAC 1200  
  
QY 1674 CACTTTTCTGTGGATAACAAAAAGGTCCTTTGGAACCTCATCTCAAGACATGGTGAA 1733  
Db 1201 CACTTTTCTGTGGATAACAAAAAGGTCCTTTGGAACCTCATCTCAAGACATGGTGAA 1260  
  
QY 1734 AGATGCACTTCGGTCTTTTGTGATCAGATCCTCCGAGTGTCTTTCCCCCAAGTGCTACCTATAT 1793  
Db 1261 AGATGCACTTCGGTCTTTTGTGATCAGTCTCTCCG--GTGCTTTCCCCCAAGTGCTACCTATAT 1318  
  
QY 1794 AATAATCACAGGCTAAGGACTGATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTTC 1853  
Db 1319 AATAATCACAGGCTAAGGACTGATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTTC 1378  
  
QY 1854 TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1913  
Db 1379 TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1438  
  
QY 1914 ATTCCTGAGGAATTTGCCACCTTCGAGGATGAGTTTATGACATTTTATTTAATAGGCAG 1973  
Db 1439 ATTCCTGAGGAATTTGCCACCTTCGAGGATG-----AGGCAG 1475  
  
QY 1974 AGAAGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 2033  
Db 1476 AGAAGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 1535  
  
QY 2034 TGGCCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGATTTATGATACAGTACCTTC 2093  
Db 1536 TGGCCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGATTTATGATACAGTACCTTC 1595  
  
QY 2094 TAAGTGGCTTTGAATTTGAACCTTACAGTATGCACGAGTACTATTACATATATTGGTATC 2153









Db 2344 TTCAGTATCAGTACAACACCCCTTACCTGGAGAGAACG-AAACTGCTGTTTAAAGTGGTT 2402  
QY 2946 TATTATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATTGATGAACGTTATATGTTT 3005  
Db 2403 TGTATACTCCATGGGTGTGACGGG----CTGGAATGCATTGGTGAACGTTACATGGTTT 2458  
QY 3006 TATTACAGATTTAATCACAATAATCATTTTTTATGAATGATTGAGTGAAAAATAGTGTTTATA 3065  
Db 2459 TATTACAGACTTCATTGTAAACTTTTTTAATGAATGATTGTGTG-AAATAGTGTTTGTA 2517  
QY 3066 AAGGTTAATAAATTTCTTGACAAAA 3091  
Db 2518 AAGGTTAATAAATTTCTTGACAAAA 2543

RESULT 7

ABS51309

ID ABS51309 standard; cDNA; 719 BP.

XX

AC ABS51309;

XX

DT 21-OCT-2002 (first entry)

XX

DE cDNA encoding human secretory protein #7.

XX

KW Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;  
KW bursitis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;  
KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;  
KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;  
KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;  
KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;  
KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;  
KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;  
KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;  
KW parkinson's disease; central nervous system disorder; mental disorder;  
KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;  
KW gene therapy; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200257304-A2.

XX

PD 25-JUL-2002.

XX

PF 15-JAN-2002; 2002WO-US001340.

XX

PR 16-JAN-2001; 2001US-0261864P.

PR 16-JAN-2001; 2001US-0261865P.

PR 16-JAN-2001; 2001US-0261979P.

PR 16-JAN-2001; 2001US-0261981P.

PR 17-JAN-2001; 2001US-0262164P.

PR 17-JAN-2001; 2001US-0262208P.

PR 17-JAN-2001; 2001US-0263131P.

PR 19-JAN-2001; 2001US-0262599P.

PR 19-JAN-2001; 2001US-0262760P.

PR 19-JAN-2001; 2001US-0263063P.

PR 19-JAN-2001; 2001US-0263066P.

PR 19-JAN-2001; 2001US-0263069P.

PR 19-JAN-2001; 2001US-0263070P.

PR 19-JAN-2001; 2001US-0263074P.

PR 19-JAN-2001; 2001US-0263076P.

PR 19-JAN-2001; 2001US-0263077P.

PR 19-JAN-2001; 2001US-0263329P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;

PI Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;

PI Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;

XX

DR WPI; 2002-590716/63.

DR P-PSDB; ABG69816.

DR

XX New purified secretory polypeptides and polynucleotides, useful in the  
PT diagnosis, study, prevention or treatment of diseases associated with  
PT decreased expression of functional secretory molecules, e.g. AIDS, cancer  
PT or allergies.  
XX Claim 1; Page 247; 340pp; English.  
PS

XX The invention describes an isolated polynucleotide a naturally occurring  
CC polynucleotide sequence at least 90 % identical to it, a polynucleotide  
CC complementary to it or an RNA equivalent of it. The purified secretory  
CC polypeptides (SPTM) and polynucleotides are useful in the diagnosis,  
CC study, prevention or treatment of diseases associated with decreased  
CC expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,  
CC bursitis, cirrhosis, hepatitis, polycythaemia vera, primary  
CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,  
CC leukaemia, myeloma or sarcoma, immune system disorder such as acquired  
CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,  
CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,  
CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,  
CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid  
CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's  
CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's  
CC disease, other developmental disorder of the central nervous system,  
CC mental disorder including mood, anxiety or schizophrenic disorder,  
CC amnesia or Tourette's disorder. The polynucleotides may be used in  
CC hybridisation and amplification technologies, e.g. in assessing gene  
CC expression patterns, to develop a transcript image for a particular cell  
CC or tissue, or to create transgenic animals to model human disease. This  
CC sequence encodes a human secretory protein isolated in the invention  
XX  
SQ Sequence 719 BP; 223 A; 113 C; 188 G; 195 T; 0 U; 0 Other;

Query Match 23.2%; Score 717.4; DB 6; Length 719;  
Best Local Similarity 99.9%; Pred. No. 5.8e-171;  
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 474 GCGCGCGCGGAGCGGCGTCTGTTATTTCCGTGGTCCGGACAGTCGCGCGCGGGT 533  
Db 1 GCGCGCGCGGAGCGGCGTCTGTTATTTCCGTGGTCCGGACAGTCGCGCGCGGGT 60  
QY 534 GACCACGGGAGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTAC 593  
Db 61 GACCACGGGAGAAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTAC 120  
QY 594 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 653  
Db 121 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTAC 180  
QY 654 CCAAGATTTTGAAGAAGCTTGTGCGAATTAAGTTGGGAGACTACTTTCATGATAAGCT 713  
Db 181 CCAAGATTTTGAAGAAGCTTGTGCGAATTAAGTTGGGAGACTACTTTCATGATAAGCT 240  
QY 714 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 773  
Db 241 ATTTGGTCTTTTGAAGCCATGCTGCTATTGAAATGATGGATCCCAAGATGGATGCTGG 300  
QY 774 CATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 833  
Db 301 CATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG 360  
QY 834 CACTATTAAAAATTAAAGATCTCACCTTGCCTGAACTGATAGGATTATGGATACATGTTT 893  
Db 361 CACTATTAAAAATTAAAGATCTCACCTTGCCTGAACTGATAGGATTATGGATACATGTTT 420  
QY 894 TTGCTGTTTGATAAACGTTGGTTAGAAAGCCATTCACTGGCACAGACAGTATTACGTGCCT 953  
Db 421 TTGCTGTTTGATAAACGTTGGTTAGAAAGCCATTCACTGGCACAGACAGTATTACGTGCCT 480  
QY 954 TTACATTTCATAAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTCTCTGGGAAT 1013  
Db 481 TTACATTTCATAAATCCAGACTTTTATAGAAGATCCTGCTATGAAGGCTTTTCTCTGGGAAT 540

QY 1014 CTTGAAATCTGTGACATTGCAAGGGAAGTAAATAAGCTGCTGTTTTTGAAGAGGA 1073  
|||||  
Db 541 CTTGAAATCTGTGACATTGCAAGGGAAGTAAATAAGCTGCTGTTTTTGAAGAGGA 600  
QY 1074 AGATTTTCAGTCAATGACTTATGGATTAAATGGCTAACAGTGTACAGATCTTCGAGT 1133  
|||||  
Db 601 AGATTTTCAGTCAATGACTTATGGATTAAATGGCTAACAGTGTACAGATCTTCGAGT 660  
QY 1134 TACAGGCATGCTAAAAAGATGTGGAGGATGACATGCAAGAAAGAGTAAAGAGTACTCGAA 1192  
|||||  
Db 661 TACAGGCATGCTAAAAAGATGTGGGGGATGACATGCAAGAAAGAGTAAAGAGTACTCGAA 719

RESULT 8  
ACH333702  
ID ACH333702 standard; cdNA; 482 BP.  
XX  
AC ACH333702;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
DE Human endothelial cell cdNA #1835.  
XX  
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2003073623-A1.  
XX  
PD 17-APR-2003.  
XX  
PF 30-JUL-2001; 2001US-00918995.  
XX  
PR 30-JUL-2001; 2001US-00918995.  
XX

(DRMA/) DRMANAC R T.  
PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX  
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX  
DR WPI; 2003-615964/58.  
XX  
PT New polynucleotide sequences obtained from various cdNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
PS Claim 1; SEQ ID NO 20914; 44pp; English.  
XX

CC The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cdNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversities, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cdNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
SQ Sequence 482 BP; 145 A; 82 C; 124 G; 130 T; 0 U; 1 Other;

Query Match 14.6%; Score 451.2; DB 8; Length 482;  
Best Local Similarity 99.3%; Pred. No. 1.1e-103;  
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 491 CGTCGTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGCGGTGACACCGGAGAAGTAG 550  
|||  
Db 26 CGAGCTTATTTCCGTGGTCCGGACAGTCCGTGGCGCGCGGTGACACCGGAGAAGTAG 85  
QY 551 GCATAATGGTTATGAAAAGCTTCTGTAGATGATGACGATTACAGATGGGAGCTCAGTATGC 610  
|||  
Db 86 GCATAATGGTTATGAAAAGCTTCTGTAGATGATGACGATTACAGATGGGAGCTCAGTATGC 145  
QY 611 CAGAAAAAATGGAGAAAAGCAATACAAAACCTGGGTGGACATTACCCAAAGATTTTGAAGAAG 670  
|||  
Db 146 CAGAAAAAATGGAGAAAAGCAATACAAAACCTGGGTGGACATTACCCAAAGATTTTGAAGAAG 205  
QY 671 CTTGTCGAGAATTAAAGCTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAG 730  
|||  
Db 206 CTTGTCGAGAATTAAAGCTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTGAAG 265  
QY 731 CCATGCTCTGCTATTGAAATGATCGATCCCAAGATGGATGCTGGCATGATTGGAACCCAAG 790  
|||  
Db 266 CCATGCTCTGCTATTGAAATGATCGATCCCAAGATGGATGCTGGCATGATTGGAACCCAAG 325  
QY 791 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAAAATTAAAG 850  
|||  
Db 326 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAAAATTAAAG 385  
QY 851 ATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTTTGTGTTGATAACGT 910  
|||  
Db 386 ATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTTTGTGTTGATAACGT 445  
QY 911 GGTAGAAGGCCATTCACTGCGCACAGACAGTATTTA 946  
|||  
Db 446 GGTAGAAGGCCATTCACTGCGCACAGACAGTATTTA 481

RESULT 9  
ADC31969  
ID ADC31969 standard; cdNA; 482 BP.  
XX  
AC ADC31969;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human novel cdNA contig sequence, SEQ ID NO:2051.  
XX  
KW Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 9; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029271-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 24-SEP-2002; 2002WO-US030474.  
XX  
PR 24-SEP-2001; 2001US-0324631P.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX  
DR WPI; 2003-371981/35.





```
CC seqdata.uspto.gov/sequence.html?DocID=20020137160
XX
SQ Sequence 398 BP; 131 A; 89 C; 85 G; 93 T; 0 U; 0 Other;

  Query Match      11.3%; Score 350.8; DB 7; Length 398;
  Best Local Similarity 94.3%; Pred. No. 2.7e-78;
  Matches 364; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2298 GCCGAGAGATCACAAATGAGCCAAAGCATATATCAGAAACATGTGTGCTGGAATGTTTAAACCA 2357
Db |||||||
QY 2358 TGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAAC 2417
Db |||||||
QY 2418 AAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCAGTGC 2477
Db 133 AAGTTCGATATGAGCACAGATTGTCTCCATTCAACAGTGTATGACACACCACCAAGTGC 192
QY 2478 ACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTC 2537
Db |||||
QY 193 ACTATCTGCAGTTCAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCAGTCTC 252
QY 2538 CTGAACTGTATGTGGCAGCTAGTAAGCACTTTCACACAGGCAAAATGATATTGGAAAATA 2597
Db |||||||
QY 253 CAGAACTGTATGTGGCAGCTAGTAAGCACTTTCAGCAGGCAAAATGATACTGGAAAATA 312
QY 2598 TTCCTAACCCGACCACATGAGGTTAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTGTGG 2657
Db |||||||
QY 313 TCCCAAAACCAGACCACATGAGGTCAATAGAAATTTTAAAGGTTGCCAAACCCCAACTTTGTGG 372
QY 2658 TTATGAAGTTATTGGCAGGAGGACAC 2683
Db |||||||
QY 373 TTATGAAGTTATTGGCAGGAGGACAC 398

RESULT 11
ABQ51659/c
ID ABQ51659 standard; DNA; 799 BP.
XX
AC ABQ51659;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38250.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
```

```
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 799 BP; 289 A; 297 C; 95 G; 118 T; 0 U; 0 Other;

  Query Match      11.1%; Score 343.8; DB 6; Length 799;
  Best Local Similarity 83.5%; Pred. No. 2.3e-76;
  Matches 390; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 87 TGCTCGTTCGTGCTCGCAACCACTAAGGTCTACGAAACCTCACCGTTTCCTTCGGCCT 146
Db |||||||
QY 798 TGTTTCGTTCGTGTTTCGTAATTATTAAAGGTTTACGTAATTTTACGGTTTTCGTTT 739
QY 147 TCGGTCACCTTTCTAAGAAATTCACAGAGGGCAGCGACAGGGGCGGGCTCTGAGACT 206
Db |||||||
QY 738 TCGCGTTATTTTAAAGAAATTTTAGAGGGTAGCGTAGACGGGCGGGTTTGTAGATT 679
QY 207 CCGGGCTCCGCTCTTTCCGGGAACCGCCCACTACCCAGGACTCCGACAGAGGGTGAATA 266
Db |||||||
QY 678 TCGGGTTCGTTTTCGGGAATCGTTTATTATTAGGATTCGATAGAGGTGAATA 619
QY 267 AAGATAAATTCGGTCTCGCGATCGTCTCTAATCTCGCGAGAAGAGAGCGCGCCCAT 326
Db |||||||
QY 618 AAGATAATTTTCGGTTCGCGATCGTTTAAATTCGCGAGAAGAGAGCGGTCGTTAT 559
QY 327 CGGCCGAACGAGCGGTGGCGAGGGGGGTGTGGCCGGGAGCGCGAAGTCCCCGGG 386
Db |||||||
QY 558 CGGTCGAACGAGCGGTGGCGAGGGAGGGGTGTGGTCGGGAGCGCGAAGTTTCGGG 499
QY 387 AGTAAGGAGAGGGGGGTTCGCGCGTCCCGGCATACGCATCGCTGCACGCTGCCGG 446
Db |||||||
QY 498 AGTAAGGAGAGGGGGGTTCGCGCGTTCGCGGTATACGTATCGGTGTACGTTGTGG 439
QY 447 TCGGCTGGGCTGAGAGGGGAGGGGCGGGCGCGCGCGCGTTCGTTATTTCGCTG 506
Db |||||||
QY 438 TCGGGTTGGGTTGAGAGGGGAGGGGCGGGCGGGCGGTCGAGGCGGTCGTTATTTCG 379
QY 507 GTCGGACAGTTCGTGGCGGGCGGGGTGACCAACGGGAGAGTAGGCA 553
Db |||||||
QY 378 GTTCGATAGTTCGTGGCGCGCGGGTGATTACGGGAGAGTAGGTA 332

RESULT 12
ABQ51658
ID ABQ51658 standard; DNA; 799 BP.
XX
AC ABQ51658;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 38249.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
```

KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP010074.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
PR 05-SEP-2000; 2000DE-01044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
DR  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
SQ Sequence 799 BP; 118 A; 95 C; 297 G; 289 T; 0 U; 0 Other;  
  
Query Match 11.1%; Score 343.8; DB 6; Length 799;  
Best Local Similarity 83.5%; Pred. No. 2.3e-76;  
Matches 390; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
  
Qy 87 TGTCTCGTTCGTCTCGCAACCACTAAGGTCTACGCAACCTCCACGGTTCCTTCCGCCT 146  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
2 TGTCTCGTTCGTCTCGTAATTATTAAGGTCTACGTAAATTTTACGGTTCCTTTCGTTT 61  
  
Qy 147 TCGGTCACCTTTCTAAGAAATTCACAGAGGCGCAGACGGGCGGGCTCTGAGACT 206  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
62 TCGCGTTATTTTAAAGAAATTTTAGAGGGTAGCGTAGACGGGCGGGTTTTCGATT 121  
  
Qy 207 CCGGTCCTCCGCTCTTTCCGGGAACGCCCACTACCCAGGACTCCGACAGAGGGTGA AAA 266  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
122 TCGGGTTTCGTTTTTTTCGGGAATCGTTTATTTAGGATTCGATAGAGGGTGA AAA 181  
  
Qy 267 AAGATAACTTCGGTCTCGCGATCGTCTCTAATCTCGCGAGAGAGAGCGGCCGCAT 326  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
182 AAGATAATTTTCGGTTTCGGATCGTTTTTAAATTCGCGAGAGAGAGCGGTCGTTAT 241  
  
Qy 327 CGGCCGAACGAGCGGTGGCGAGGAGGGGGTGTGGCCCGGGAGCGCAAGTCCCCGGG 386  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
242 CGGTCGAACGAGCGGTGGCGAGGAGGGGGTGTGGTCCGGGAGCGCAAGTTTTCGGG 301

Qy 387 AGTAAGGGAGAGGGGGCGGGTCCGCGTCCCGGGCATACGATCGGTGCACGCTGCCGG 446  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
302 AGTAAGGGAGAGGGGGCGGGTCCGCGTTCGGGTATACGTATCGGTGTACGTTGTCCG 361  
  
Qy 447 TCGGCTGGGCTGAGAGGGGAGGGGGCGGGCGCGCGGCGGAGGCGCGCTCGTTATTTCCGTG 506  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
362 TCGGTTGGGTTGAGAGGGGAGGGGGCGGGCGGGTTCGAGGCGCGCTCGTTATTTTCGTG 421  
  
Qy 507 GTCGGACAGTGC GTGGCGGGCGGGTGACCA CGGGAGAGTAGGCA 553  
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
422 GTTCGATAGTGC GTGGCGGGCGGGTGATTACGGGAGAGTAGGTA 468  
  
RESULT 13  
ABV15331/c  
ID ABV15331 standard; cDNA; 356 BP.  
XX  
AC ABV15331;  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 15322.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US005171.  
XX  
PR 17-FEB-2000; 2000US-0183319P.  
PR 16-MAR-2000; 2000US-0189862P.  
PR 25-MAY-2000; 2000US-0207454P.  
PR 09-JUN-2000; 2000US-0211314P.  
PR 18-JUL-2000; 2000US-0219007P.  
PR 13-DEC-2000; 2000US-0255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 2572; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 U; 0 Other;  
  
Query Match 9.4%; Score 292.4; DB 5; Length 356;  
Best Local Similarity 90.5%; Pred. No. 1.5e-63;  
Matches 344; Conservative 0; Mismatches 11; Indels 25; Gaps 2;





XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
XX Claim 1; Page 7483; 11750pp; English.  
PS  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 U; 0 Other;  
  
Query Match 9.3%; Score 288.8; DB 5; Length 439;  
Best Local Similarity 91.4%; Pred. No. 1.4e-62;  
Matches 338; Conservative 0; Mismatches 7; Indels 25; Gaps 2;  
  
QY 1687 GATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGAATGCACCTTCGG 1746  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
439 GATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAGAATGCACCTTCGG 380  
  
QY 1747 TCTTTTGTTCAGATCCTCCGAGTGTCTTCCCAAGTGCTACCTATATAATAATCACCAGG 1806  
Db ||||||||||| | ||||||||||||||||||||||||||||||||||||||||  
379 TCTTTTGTTCAG--TCCTCCGGTGTCTTCCCAAGTGCTACCTATATAATAATCACCAGG 322  
  
QY 1807 CTAGGACTGTATCGACTCCTTTGTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTC 1866  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
321 CTAGGACTGTATCGACTCCTTTGTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTC 262  
  
QY 1867 AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAAT 1926  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
261 AGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAAT 202  
  
QY 1927 TTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGTTGATGC 1986  
Db |||||||||||||||| | |||||||||||-----AGGCAGAGAAGTTGATGC 165  
201 TTGCCACCTTGCAGGATG-----  
  
QY 1987 AGCGCTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 2046  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
164 AGCGCTTCACACCATGCTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGG 105  
  
QY 2047 TACCTGGGTC 2056  
Db ||||| | |  
104 TACCTCGGCC 95

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 21:58:32 ; Search time 1436.92 Seconds  
(without alignments)  
10571.835 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttctcacgaaactcccagg.....atttcttgacaaaaaaa 3096

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3096	100.0	3096	14	US-10-001-857-42 Sequence 42, Appl
2	2417.2	78.1	2488	17	US-10-467-433-23 Sequence 23, Appl
C 3	2228.4	72.0	2610	10	US-09-814-353-20725 Sequence 20725, A
C 4	560.8	18.1	745	10	US-09-814-353-15927 Sequence 15927, A
5	509.4	16.5	610	10	US-09-814-353-16155 Sequence 16155, A
C 6	496.2	16.0	602	10	US-09-814-353-3218 Sequence 3218, Ap
C 7	496.2	16.0	602	10	US-09-814-353-9543 Sequence 9543, Ap
8	451.2	14.6	482	10	US-09-918-995-20914 Sequence 20914, A
9	434	14.0	434	10	US-09-814-353-3455 Sequence 3455, Ap
10	434	14.0	434	10	US-09-814-353-9771 Sequence 9771, Ap
11	350.8	11.3	398	9	US-09-983-965-210 Sequence 210, App
12	302.6	9.8	345	13	US-10-085-783A-43407 Sequence 43407, A
13	302.6	9.8	345	16	US-10-242-535A-43407 Sequence 43407, A
14	268.6	8.7	501	9	US-09-783-590-5693 Sequence 5693, Ap

C 15	182	5.9	489	13	US-10-027-632-323193	Sequence 323193,
C 16	182	5.9	489	16	US-10-027-632-323193	Sequence 323193,
C 17	140.4	4.5	151	14	US-10-001-857-41	Sequence 41, Appl
C 18	87.8	2.8	815	14	US-10-001-857-111	Sequence 111, App
19	81.4	2.6	430	13	US-10-085-783A-54186	Sequence 54186, A
20	81.4	2.6	430	16	US-10-242-535A-54186	Sequence 54186, A
21	79.2	2.6	962	13	US-10-424-599-47623	Sequence 47623, A
22	64.2	2.1	2121	17	US-10-437-963-44335	Sequence 44335, A
23	60.8	2.0	456	17	US-10-437-963-51928	Sequence 51928, A
24	60.2	1.9	629	17	US-10-021-323-9375	Sequence 9375, Ap
25	60	1.9	60	10	US-09-908-975-16495	Sequence 16495, A
C 26	59.8	1.9	492	13	US-10-424-599-96349	Sequence 96349, A
C 27	59.6	1.9	852	17	US-10-437-963-68188	Sequence 68188, A
C 28	58.6	1.9	420	17	US-10-021-323-5998	Sequence 5998, Ap
29	58.2	1.9	3163	15	US-10-017-161-1857	Sequence 1857, Ap
30	58.2	1.9	3163	16	US-10-292-798-1513	Sequence 1513, Ap
C 31	58	1.9	925	17	US-10-437-963-44536	Sequence 44536, A
C 32	58	1.9	956	13	US-10-027-632-31508	Sequence 31508, A
C 33	58	1.9	956	16	US-10-027-632-31508	Sequence 31508, A
34	57.8	1.9	809	13	US-10-424-599-26988	Sequence 26988, A
C 35	57.4	1.9	497	13	US-10-424-599-133788	Sequence 133788,
C 36	56.6	1.8	2432	17	US-10-437-963-64501	Sequence 64501, A
C 37	56.4	1.8	580	13	US-10-424-599-2320	Sequence 2320, Ap
C 38	56.2	1.8	1211	17	US-10-437-963-27585	Sequence 27585, A
C 39	56.2	1.8	5452	15	US-10-017-161-1481	Sequence 1481, Ap
40	56.2	1.8	5452	16	US-10-292-798-1189	Sequence 1189, Ap
41	56	1.8	451	13	US-10-240-425-721	Sequence 721, App
C 42	56	1.8	630	17	US-10-437-963-18774	Sequence 18774, A
43	56	1.8	1117	15	US-10-017-161-1403	Sequence 1403, Ap
44	56	1.8	1117	16	US-10-292-798-1141	Sequence 1141, Ap
C 45	55.8	1.8	424	17	US-10-437-963-49515	Sequence 49515, A

ALIGNMENTS

RESULT 1

US-10-001-857-42  
; Sequence 42, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42  
; LENGTH: 3096  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-001-857-42

Query Match 100.0%; Score 3096; DB 14; Length 3096;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3096; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTCACGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 60

Db 1 TTCCTCACGAAACTCCAGGCGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 60

QY 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTGCTCGTCTCGCAACCACTAAGGTCTACG 120

Db 61 GCACGCTGTTGCTCTCGGAGCTTGGCTGCTGCTGCTCGTCTCGCAACCACTAAGGTCTACG 120

QY 121 CAAACCTCCACGGTTTCCTTCGCGCTTCGCGTCACTTTCTTAAGAAATCCCAGAGGCA 180



Db 121 CAAACCTCCACGGTTTCCTCCGCCCTTCGCGTCACCTTTCTAAGAAATCCAGAGGGCA 180  
QY 181 GCGCAGACGGGGCGGCTCTGAGACTCCGGGCTCCGCCTCTTTCCGGGAACCGCCCCACTA 240  
Db 181 GCGCAGACGGGGCGGCTCTGAGACTCCGGGCTCCGCCTCTTTCCGGGAACCGCCCCACTA 240  
QY 241 CCCAGGACTCCGACAGAGGGTGAAAAAGATAAACTTCCGGTCTCGGATCGTCTCTAATC 300  
Db 241 CCCAGGACTCCGACAGAGGGTGAAAAAGATAAACTTCCGGTCTCGGATCGTCTCTAATC 300  
QY 301 TCGCGAGAAAGAGCGCGGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGGGTG 360  
Db 301 TCGCGAGAAAGAGCGCGGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGGGTG 360  
QY 361 TGGCCGGGAGCGCGAAGTCCCGGAGTAAGGGAGAGGGGCGGGTCCGCGTCCCGG 420  
Db 361 TGGCCGGGAGCGCGAAGTCCCGGAGTAAGGGAGAGGGGCGGGTCCGCGTCCCGG 420  
QY 421 GCATACGCATCGTGCACGCTGCCGCTCGGCTGGGCTGAGAGGGGAGGGGGCGCGCG 480  
Db 421 GCATACGCATCGTGCACGCTGCCGCTCGGCTGGGCTGAGAGGGGAGGGGGCGCGCG 480  
QY 481 GCCGAGGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGCGCGCGCGGGTGACCA 540  
Db 481 GCCGAGGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGGCGCGCGCGGGTGACCA 540  
QY 541 GGAGAAGTAGGCATAATGGTTATGAAAGTTCTGTAGATGATGACATTCAGGATGGGAG 600  
Db 541 GGAGAAGTAGGCATAATGGTTATGAAAGTTCTGTAGATGATGACATTCAGGATGGGAG 600  
QY 601 CTCAGTATGCCAGAAAAAATGGAGAAAGCAATACAACTGGGTGGACATTAACCAAGAT 660  
Db 601 CTCAGTATGCCAGAAAAAATGGAGAAAGCAATACAACTGGGTGGACATTAACCAAGAT 660  
QY 661 TTTGAAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTGGT 720  
Db 661 TTTGAAAGCTTGTGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCTATTGGT 720  
QY 721 CTTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGAT 780  
Db 721 CTTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGATGCTGGCATGAT 780  
QY 781 GGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGGCACTAT 840  
Db 781 GGAACCAAGTTAATCGAAAAGTTCTCAATTTTGAAACAAGCTATCAAGGATGGCACTAT 840  
QY 841 AAAATTAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTGTCTGT 900  
Db 841 AAAATTAAAGATCTCACCTTGCCCTGAACTGATAGGGATTATGGATACATGTTTGTCTGT 900  
QY 901 TTGATAACGTGTTAGAAAGCCATTCACTGGCACAGACAGTAFTTACGTGCCCTTACAT 960  
Db 901 TTGATAACGTGTTAGAAAGCCATTCACTGGCACAGACAGTAFTTACGTGCCCTTACAT 960  
QY 961 CATATCCAGACTTTATAGAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
Db 961 CATATCCAGACTTTATAGAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAA 1020  
QY 1021 ATCTGTGACATTGCAAGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTT 1080  
Db 1021 ATCTGTGACATTGCAAGGAAAAAGTAAATAAAGCTGCTGTTTGAAGAGGAAGATTT 1080  
QY 1081 CAGTCAATGACTTATGGATTTAAATGGCTAACAGTGTGACAGATCTTCAGTTACAGGC 1140  
Db 1081 CAGTCAATGACTTATGGATTTAAATGGCTAACAGTGTGACAGATCTTCAGTTACAGGC 1140  
QY 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAGAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
Db 1141 ATGCTAAAAGATGTGGAGGATGACATGCAAGAGAGTAAAGAGTACTCGAAGTCGACAA 1200  
QY 1201 GGAGAAGAAAGATCCAGAAGTTGAACTAGAACACCAATGTTTAGCAGTATTCAGC 1260  
Db 1201 GGAGAAGAAAGATCCAGAAGTTGAACTAGAACACCAATGTTTAGCAGTATTCAGC 1260

Db 1201 GGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTCAGC 1260  
QY 1261 AGAGTGAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCCTTTACTAAGAAAGAGACC 1320  
Db 1261 AGAGTGAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCCTTTACTAAGAAAGAGACC 1320  
QY 1321 AGTCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATT 1380  
Db 1321 AGTCTGTTGCAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTGCCATT 1380  
QY 1381 CATAATTCATTCATCATGGCATCCAGGCCCAGAAATGATACATAAAAAGGAGATCATCCA 1440  
Db 1381 CATAATTCATTCATCATGGCATCCAGGCCCAGAAATGATACATAAAAAGGAGATCATCCA 1440  
QY 1441 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500  
Db 1441 ATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGA 1500  
QY 1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGCAGATTAATAGATAGA 1560  
Db 1501 TATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACTATTTTGCAGATTAATAGATAGA 1560  
QY 1561 ATAAAAACTGTCTGTGAGGTTGTGNAATTTAACAAATTTACATTTGATCTCTGGATTTTTTC 1620  
Db 1561 ATAAAAACTGTCTGTGAGGTTGTGNAATTTAACAAATTTACATTTGATCTCTGGATTTTTTC 1620  
QY 1621 TGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTACAAACCACTTTC 1680  
Db 1621 TGTGAATTTAGTGAACAGTCAACCATGTGTTCTTTCAAGATCTCTGTACAAACCACTTTC 1680  
QY 1681 CTGTGATTAACAAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA 1740  
Db 1681 CTGTGATTAACAAAAAGGTCTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCA 1740  
QY 1741 CTTCCGTCTTTTGTGATCAGATCCCTCCGAGTGTCTTCCCCCAAGTGTACCTATATAATAATC 1800  
Db 1741 CTTCCGTCTTTTGTGATCAGATCCCTCCGAGTGTCTTCCCCCAAGTGTACCTATATAATAATC 1800  
QY 1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTGGCCATTTCTGTAGTC 1860  
Db 1801 ACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTGGCCATTTCTGTAGTC 1860  
QY 1861 TTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920  
Db 1861 TTATTTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920  
QY 1921 AGGAATTTGCCACCTTGACAGATGAGTTTATGACATTTTAAATAGGCAGAGAAGT 1980  
Db 1921 AGGAATTTGCCACCTTGACAGATGAGTTTATGACATTTTAAATAGGCAGAGAAGT 1980  
QY 1981 TGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTG 2040  
Db 1981 TGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTG 2040  
QY 2041 TTTAGGTACCTGGGTCTTTTACCATAACTTCGCAATTAATGATACAGTACCTTCTAAGTGG 2100  
Db 2041 TTTAGGTACCTGGGTCTTTTACCATAACTTCGCAATTAATGATACAGTACCTTCTAAGTGG 2100  
QY 2101 CTTTGAATTGGAACTCTACAGTATGCACGAGTACTATTAATATATATGGTATCTCTCTGA 2160  
Db 2101 CTTTGAATTGGAACTCTACAGTATGCACGAGTACTATTAATATATATGGTATCTCTCTGA 2160  
QY 2161 ATTCCTTTACGATGTTGATGTCAACATTTGAGTCGTGCGGATGGCTCTCAAATGGCAGA 2220  
Db 2161 ATTCCTTTACGATGTTGATGTCAACATTTGAGTCGTGCGGATGGCTCTCAAATGGCAGA 2220  
QY 2221 GGAAGGATTAATGGAAGAGCAGAGAAAGGCCGTAGTAAATAAAACAAAGAAAAA 2280  
Db 2221 GGAAGGATTAATGGAAGAGCAGAGAAAGGCCGTAGTAAATAAAACAAAGAAAAA 2280  
QY 2281 GAAAGTTCCGCCATTGAGCCGAGAGATCAATGAGCCAGCATATCAGAACATGTGTGC 2340  
Db 2281 GAAAGTTCCGCCATTGAGCCGAGAGATCAATGAGCCAGCATATCAGAACATGTGTGC 2340

Qy 2341 TGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTT 2400  
Db |||||||  
Qy 2341 TGAATGTTTAAACCATGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTT 2400  
Db |||||||  
Qy 2401 TGAGCTTGATAGTGAACAAGTTCCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGAT 2460  
Db |||||||  
Qy 2401 TGAGCTTGATAGTGAACAAGTTCCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGAT 2460  
Db |||||||  
Qy 2461 GACCCCGCGCAGTGCACTACTTACAGTTCAAGGAAATGCTGACCTCAATAAATATAG 2520  
Db |||||||  
Qy 2461 GACCCCGCGCAGTGCACTACTTACAGTTCAAGGAAATGCTGACCTCAATAAATATAG 2520  
Db |||||||  
Qy 2521 CCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
Db |||||||  
Qy 2521 CCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAA 2580  
Db |||||||  
Qy 2581 AATGATATTGGAATAATTTCCTAACCCGACCATGAGGTTAATAGAAATTTAAAGGTTGC 2640  
Db |||||||  
Qy 2581 AATGATATTGGAATAATTTCCTAACCCGACCATGAGGTTAATAGAAATTTAAAGGTTGC 2640  
Db |||||||  
Qy 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
Db |||||||  
Qy 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAATCTAAAGT 2700  
Db |||||||  
Qy 2701 TCCTCCTGAATTTGAATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTGTTTGAGA 2760  
Db |||||||  
Qy 2701 TCCTCCTGAATTTGAATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTGTTTGAGA 2760  
Db |||||||  
Qy 2761 GAGACTGGGAGGTGGCCATAAAGGGGAGGCTCTTTTTCAGACCCCACTTTAGAGGG 2820  
Db |||||||  
Qy 2761 GAGACTGGGAGGTGGCCATAAAGGGGAGGCTCTTTTTCAGACCCCACTTTAGAGGG 2820  
Db |||||||  
Qy 2821 CACATCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACTCATT 2880  
Db |||||||  
Qy 2821 CACATCACCAGGCTCCACATCACGGGAAGTGAGATGGATTTCTTGGGTAACAACTCATT 2880  
Db |||||||  
Qy 2881 TAAGGAATACTTTTAGTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940  
Db |||||||  
Qy 2881 TAAGGAATACTTTTAGTTGACAGCCTTATATGACATGAATGAAACTGCTGTTTAAAG 2940  
Db |||||||  
Qy 2941 TGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATAT 3000  
Db |||||||  
Qy 2941 TGGTTTATTATGTTCCATGGAAGAACTGGTCTTATTGAATGCATTGATGAACGTTATAT 3000  
Db |||||||  
Qy 3001 GGTTTTATTACAGATTTAATCACAATCATTTTTTATGAATGATTGAGTGAAAATAGTGT 3060  
Db |||||||  
Qy 3001 GGTTTTATTACAGATTTAATCACAATCATTTTTTATGAATGATTGAGTGAAAATAGTGT 3060  
Db |||||||  
Qy 3061 TTATAAAGGTTAATAAAATTTCTTGACAAAAA 3096  
Db |||||||  
Qy 3061 TTATAAAGGTTAATAAAATTTCTTGACAAAAA 3096  
Db |||||||

RESULT 2

US-10-467-433-23  
; Sequence 23, Application US/10467433  
; Publication No. US2004008773A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION;  
; APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;  
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;  
; APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;  
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;  
; APPLICANT: DING, Li; GIETZEN, Kimberly J.;  
; APPLICANT: ISON, Craig H.; LU, Dyung Aina M.;  
; APPLICANT: HAFALIA, April J.A.; GANDHI, Ameena R.;  
; APPLICANT: THANGAVELU, Kavitha; SANJANWALA, Madhusudan M.;  
; APPLICANT: TANG, Y. Tom; RAMKUMAR, Jayalaxmi;  
; APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;  
; APPLICANT: AZIMZAI, Yalda; SAPPERSTEIN, Stephanie K.;  
; APPLICANT: BURFORD, Neil; LEE, Ernestine A.;  
; APPLICANT: LU, Yan; TRAN, Uyen K.;

; APPLICANT: MARQUIS, Joseph P.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
; FILE REFERENCE: PF-0899 USN  
; CURRENT APPLICATION NUMBER: US/10/467,433  
; CURRENT FILING DATE: 2003-08-06  
; PRIOR APPLICATION NUMBER: PCT/US02/03709  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/268,117  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/269,618  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: US 60/271,118  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: US 60/274,486  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/274,436  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: US 60/334,229  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: US 60/353,284  
; PRIOR FILING DATE: 2002-02-01  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PERL Program  
; SEQ ID NO 23  
; LENGTH: 2488  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 7111920CB1  
US-10-467-433-23

Query Match 78.1%; Score 2417.2; DB 17; Length 2488;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 2470; Conservative 0; Mismatches 13; Indels 25; Gaps 2;

Qy 474 GCGGGCGCGGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGCGGT 533  
Db ||||||| 1 GCGGGCGCGGAGCGCGTCTGTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGCGGT 60  
Qy 534 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 593  
Db ||||||| 61 GACCACGGGAGAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGG 120  
Qy 594 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAACTGGGTGGACATTAC 653  
Db ||||||| 121 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAACTGGGTGGACATTAC 180  
Qy 654 CCAAGATTTTGAAGAGCTTGTTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 713  
Db ||||||| 181 CCAAGATTTTGAAGAGCTTGTTCGAGAAATTAAGTTGGGAGAACTACTTCATGATAAGCT 240  
Qy 714 ATTTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATCTGG 773  
Db ||||||| 241 ATTTGGTCTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATCTGG 300  
Qy 774 CATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGG 833  
Db ||||||| 301 CATGATTGGAACCAAGTTAATCGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGG 360  
Qy 834 CACTATTAAAATTAAGATCTCACCTTGCCCTGAACCTGATAGGGATTATGGATACATGTTT 893  
Db ||||||| 361 CACTATTAAAATTAAGATCTCACCTTGCCCTGAACCTGATAGGGATTATGGATACATGTTT 420  
Qy 894 TTGCTGTTTATGATAACGTTGTTAGAAAGGCCATTACCTGGCACAGACAGTATTTACGTCCT 953  
Db ||||||| 421 TTGCTGTTTATGATAACGTTGTTAGAAAGGCCATTACCTGGCACAGACAGTATTTACGTCCT 480  
Qy 954 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGCTTTTGTCTGGGAAT 1013  
Db ||||||| 481 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGCTTTTGTCTGGGAAT 540  
Qy 1014 CTTGAAAATCTGTGACATTGCAAGGAAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGA 1073



||||| 541 CTTGAAAAATCTGTGACATTGCAAGGGAAGTAAATAAGCTGCTGTTTTTGAAGAGGA 600  
QY 1074 AGATTTTCAGTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133  
Db 601 AGATTTTCAGTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGT 660  
QY 1134 TACAGGCATGCTAAAAAGATGTGAGGATGACATGCAAAAGAGTAAAGAGTACTCGAAG 1193  
Db 661 TACAGGCATGCTAAAAAGATGTGAGGATGACATGCAAAAGAGTAAAGAGTACTCGAAG 720  
QY 1194 TCGACAAGGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGT 1253  
Db 721 TCGACAAGGAGAAAGAGATCCAGAAAGTTGAACTAGAACACCAACAATGTTTAGCAGT 780  
QY 1254 ATTCAGCAGAGTGAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAA 1313  
Db 781 ATTCAGCAGAGTGAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTAAGAA 840  
QY 1314 AGAGACCAGTGTGTCAGAAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTC 1373  
Db 841 AGAGACCAGTGTGTCAGAAAGCTCAAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTC 900  
QY 1374 TGCCATTTCATAATTCATTCATGCGCATCCAGSCCCAGAATGATACATAAAAGGAGA 1433  
Db 901 TGCCATTTCATAATTCATTCATGCGCATCCAGSCCCAGAATGATACATAAAAGGAGA 960  
QY 1434 TCATCCAATTATGATGGGTTTTTGAAACCCCTTGTAACCAAGGCTACTTCCACCTACCTT 1493  
Db 961 TCATCCAATTATGATGGGTTTTTGAAACCCCTTGTAACCAAGGCTACTTCCACCTACCTT 1020  
QY 1494 CCCTCGATATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGCAGATTAAT 1553  
Db 1021 CCCTCGATATGCAAAAAATAATTAAAGGGAAGAAATGGTGAACCTATTTTGCAGATTAAT 1080  
QY 1554 AGATAGATAAAAACTGTCTGTGAGGTTGTGAAATTTAACAAATTTACATTTGTAACAA 1613  
Db 1081 AGATAGATAAAAACTGTCTGTGAGGTTGTGAAATTTAACAAATTTACATTTGTAACAA 1140  
QY 1614 TTTTCTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAAC 1673  
Db 1141 TTTTCTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTACAAAC 1200  
QY 1674 CACTTTCCTGGTGATAACAAAAAGGTCCTTTGGAACTCATCTCATGCAAGACATGGTGAA 1733  
Db 1201 CACTTTCCTGGTGATAACAAAAAGGTCCTTTGGAACTCATCTCATGCAAGACATGGTGAA 1260  
QY 1734 AGATGCACTTCGGTCTTTTGTGATCCTCCGAGTGTCTTTCCCAAGTGTCTACCTATAT 1793  
Db 1261 AGATGCACTTCGGTCTTTTGTGATCCTCCGAGTGTCTTTCCCAAGTGTCTACCTATAT 1318  
QY 1794 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTC 1853  
Db 1319 AATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTTCGGCCATTC 1378  
QY 1854 TGTAGTCTTATTCAGATCCATGGACATAAACAGGGCTCGACAGAGAGATAAGCTTGGTCAAT 1913  
Db 1379 TGTAGTCTTATTCAGATCCATGGACATAAACAGGGCTCGACAGAGAGATAAGCTTGGTCAAT 1438  
QY 1914 ATTCTTGAGGAATTTGCCACCTTGACGATGAGTTTATGACATTTTATTTAAATAGGCAG 1973  
Db 1439 ATTCTTGAGGAATTTGCCACCTTGACGATG-----AGGCAG 1475  
QY 1974 AGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 2033  
Db 1476 AGAAGGTTGATGCAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAGGCAACATT 1535  
QY 2034 TGGCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTC 2093  
Db 1536 TGGCTGTTTAGGTACCTGGGTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTC 1595  
QY 2094 TAACTGGCTTTGAATTGGAACCTACAGTATGCACGAGTACTATTACATATATTGGTATC 2153  
|||||

Db 1596 TAACTGGCTTTGAATTTGGAACCTCTACAGTATGCACGAGTACTATTACATATATTGGTATC 1655  
QY 2154 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTTAGTGGTCCGATGGCTCTCAAA 2213  
Db 1656 TCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTTAGTGGTCCGATGGCTCTCAAA 1715  
QY 2214 TGGCAGAGGAAAGGATTAATGGAAGAGCAGCAGAAAAGGCGTAGTAGTAAAAAACCAAGA 2273  
Db 1716 TGGCAGAGGAAAGGATTAATGGAAGAGCAGCAGAAAAGGCGTAGTAGTAAAAAACCAAGA 1775  
QY 2274 AAAAAAGAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACA 2333  
Db 1776 AAAAAAGAAAGTTTCGCCATTGAGCCGAGAGATCACAATGAGCCAAAGCATATCAGAACA 1835  
QY 2334 TGTGTGCTGGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAC 2393  
Db 1836 TGTGTGCTGGAATGTTTAAAAACCATGGTAGCATTTGACATGGACGGCAAAAGTACGTAAC 1895  
QY 2394 CGAAGTTTGAGCTTTGATAGTGAAACAAGTTTCGGTATGAACACACAGGTTTGCTCCATTCAACA 2453  
Db 1896 CGAAGTTTGAGCTTTGATAGTGAAACAAGTTTCGGTATGAACACACAGGTTTGCTCCATTCAACA 1955  
QY 2454 GTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA 2513  
Db 1956 GTGTGATGACCCCGCCGAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATA 2015  
QY 2514 AATATAGCCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAAC 2573  
Db 2016 AATATAGCCCTCCTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAAC 2075  
QY 2574 AGGCAAAATGATATTTGAAAAATATTCCTAACCCCGACCATGAGGTTAATAGAAATTTTAA 2633  
Db 2076 AGGCAAAATGATATTTGAAAAATATTCCTAACCCCGACCATGAGGTTAATAGAAATTTTAA 2135  
QY 2634 AGGTTGCCAAACCCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT 2693  
Db 2136 AGGTTGCCAAACCCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAAT 2195  
QY 2694 CTAAAGTTCTCCTCCTGAAATTTTGATTTTCTCTGCTCATAAATATTTTCCCTGTGTGAAACTTG 2753  
Db 2196 CTAAAGTTCTCCTCCTGAAATTTTGATTTTCTGCTCATAAATATTTTCCCTGTGTGAAACTTG 2255  
QY 2754 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTTTTCAGACCCCAACTCT 2813  
Db 2256 TTTGAGAGAGACTGGGAGGTGGCCATAAAGGGGCAGAGTCTTTTTCAGACCCCAACTCT 2315  
QY 2814 TAGAGGCACATCACAGGCTCCACATCACGGGAAGTGAATGGAATTTCTTGGGTAACAA 2873  
Db 2316 TAGAGGCACATCACAGGCTCCACATCACGGGAAGTGAATGGAATTTCTTGGGTAACAA 2375  
QY 2874 CTCAATTAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACCTGCTGT 2933  
Db 2376 CTCAATTAAGGAATACCTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACCTGCTGT 2435  
QY 2934 TTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAAT 2981  
Db 2436 TTTAAAGTGGTTTATTATGTTCCATGTAAGACACCTGGGTTCCATTAAAT 2483  
|||||

RESULT 3

US-09-814-353-20725/c  
; Sequence 20725, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21



; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20725
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20725

	Query Match	72.0%;	Score 2228.4;	DB 10;	Length 2610;	
	Best Local Similarity	97.9%;	Pred. No. 0;			
	Matches 2299;	Conservative	0;	Mismatches	21;	Indels 28; Gaps 3;
QY	426	CGCATGCGTGACGCTGCCGTC	CGGTCGGGCTGGGTGAGAGGGAGGGGCGGCGGCCGA	485		
Db	2610	CGCATGCGTGACGCTGCCGTC	CGGTCGGGTGAGAGGGAGGGGCGGCGGCCGA	2551		
QY	486	GGCGGCGTCGTTATTTCCG	TGGTCCGGACAGTCGCTGGCGCGCGGTCACCGGGAGA	545		
Db	2550	GGCGGCGTCGTTATTTCCG	TGGTCCGGACAGTCGCTGGCGCGCGGTCACCGGGAGA	2491		
QY	546	AGTAGGCATAATGGTTATG	AAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAG	605		
Db	2490	AGTAGGCATAATGGTTATG	AAAGCTTCTGTAGATGATGACGATTCAGGATGGAGCTCAG	2431		
QY	606	TATGCCAGAAAAAATGG	AGAAAAGCAATACAACTGGGTGGACATTACCCAAAGATTTGA	665		
Db	2430	TATGCCAGAAAAAATGG	AGAAAAGCAATACAACTGGGTGGACATTACCCAAAGATTTGA	2371		
QY	666	AGAAGCTTGTGAGAAATT	AAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTT	725		
Db	2370	AGAAGCTTGTGAGAAATT	AAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTT	2311		
QY	726	TGAAGCCATGTCGCTATT	GAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGA	785		
Db	2310	TGAAGCCATGTCGCTATT	GAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGA	2251		
QY	786	CCAAGTTAATCGAAAAAG	TTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAT	845		
Db	2250	CCAAGTTAATCGAAAAAG	TTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAT	2191		
QY	846	TAAAGATCTCACCTTGCC	TGAAGTATAGGATTTATGGATACATGTTTTTGTGTTGAT	905		
Db	2190	TAAAGATCTCACCTTGCC	TGAAGTATAGGATTTATGGATACATGTTTTTGTGTTGAT	2131		
QY	906	AACGTGGTTAGAGGCCAT	TCACTGGCACAGACAGTATTACGTGCCCTTACATTCATA	965		
Db	2130	AACGTGGTTAGAGGCCAT	TCACTGGCACAGACAGTATTACGTGCCCTTACATTCATA	2071		
QY	966	TCCAGACTTTATAGAAG	ATCCTGCTATGAAGCTTTTGCTCTGGGAATCTGAAAAATCTG	1025		
Db	2070	TCCAGACTTTATAGAAG	ATCCTGCTATGAAGCTTTTGCTCTGGGAATCTGAAAAATCTG	2011		
QY	1026	TGACATTGCAAGGAAAAA	GTAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC	1085		
Db	2010	TGACATTGCAAGGAAAAA	GTAAAGCTGCTGTTTTTGAAGAGGAAGATTTTCAGTC	1951		
QY	1086	AATGACTTATGGAATTT	AAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT	1145		

Db	1950	AATGACTTATGGATT	TTAAAATGGCTAAACAGTGTGACAGATCTTCGAGTTACAGGCATGCT	1891		
QY	1146	AAAAGATGTGGAGGATG	ACATGCAAAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGA	1205		
Db	1890	AAAAGATGTGGAGGATG	ACATGCAAAAGAAGAGTAAAGAGTACTCGAAGTCGACAAGGAGA	1831		
QY	1206	AGAAAGAGATCCAGAAG	TTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAGT	1265		
Db	1830	AGAAAGAGATCCAGAAG	TTGAACCTAGAACACCAACAATGTTTAGCAGTATTCAGCAGAGT	1771		
QY	1266	GAAATTTACTCGTGTG	TTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACCAGTGC	1325		
Db	1770	GAAATTTACTCGTGTG	TTACTGACAGTGTCTTATAGCCTTTTACTAAGAAAAGAGACCAGTGC	1711		
QY	1326	TGTTGCAGAAAGCTCA	AAAAATTGATGTTCAAGCAGCAGATCTTCTTCTGCCATTCAATA	1385		
Db	1710	TGTTGCAGAAAGCTCA	AAAAATTGATGTTCAAGCAGCAGATCTTCTTCTGCCATTCAATA	1651		
QY	1386	TTCAATTGCATCATGG	CATCCAGGCCCAAGATGATACTACAAAAGGAGATCATCCAATTAT	1445		
Db	1650	TTCAATTGCATCATGG	CATCCAGGCCCAAGATGATACTACAAAAGGAGATCATCCAATTAT	1591		
QY	1446	GATGGGTTTTGAACCC	CTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC	1505		
Db	1590	GATGGGTTTTGAACCC	CTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC	1531		
QY	1506	AAAAATAATTAAAAGG	GAAGAAATGGTGAACCTATTTTGCAAGATTAAATAGATAGAATAAA	1565		
Db	1530	AAAAATAATTAAAAGG	GAAGAAATGGTGAACCTATTTTGCAAGATTAAATAGATAGAATAAA	1471		
QY	1566	AACTGTCGTGAGGTTG	TGAATTTAACAAAATTTACATTGTATCCTGGATTTTCTGTGA	1625		
Db	1470	AACTGTCGTGAGGTTG	TGAATTTAACAAAATTTACATTGTATCCTGGATTTTCTGTGA	1411		
QY	1626	ATTTAGTGAACAGTCA	CCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGT	1685		
Db	1410	ATTTAGTGAACAGTCA	CCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGT	1351		
QY	1686	GGATAACAAAAAGGT	CTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCG	1745		
Db	1350	GGATAACAAAAAGGT	CTTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCG	1291		
QY	1746	GTCTTTTGTGAGATCC	TCCGAGTGCCTTCCCCCAAGTGCTACCTATATAATAATCACCAG	1805		
Db	1290	GTCTTTTGTGAGATCC	TCCGAGTGCCTTCCCCCAAGTGCTACCTATATAATAATCACCAG	1233		
QY	1806	GCTAAGGACTGTATCG	ACTCCTTTGTACTCACTGTGTTCCGCCATCTCTAGTCTTATT	1865		
Db	1232	GCTAAGGACTGTATCG	ACTCCTTTGTACTCACTGTGTTCCGCCATCTCTAGTCTTATT	1173		
QY	1866	CAGATCCATGACATAA	CAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA	1925		
Db	1172	CAGATCCATGACATAA	CAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA	1113		
QY	1926	TTTGCCACCTTGCAGG	ATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGGTTGATG	1985		
Db	1112	TTTGCCACCTTGCAGG	ATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGGTTGATG	1076		
QY	1986	CAGCGCTTCACACCAT	GTCGTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAG	2045		
Db	1075	CAGCGCTTCACACCAT	GTCGTGTTGAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAG	1016		
QY	2046	GTACCTGGTTCCTTTA	CCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG	2105		
Db	1015	GTACCTGGTTCCTTTA	CCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGGCTTTG	956		
QY	2106	AATTGGAACTCTACAG	TATGCACGAGTACTATTACATATATTTGGTATCTCTCTGAATTCC	2165		
Db	955	AATTGGAACTCTACAG	TATGCACGAGTACTATTACATATATTTGGTATCTCTCTGAATTCC	896		
QY	2166	TTTACGCATGGTTGAT	GTCAACATTGAGTCGTGCCGATGGCTCTCAAATGGCAGAGGAAA	2225		

Db 895 TTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTCTCAAATGCCAGAGAAA 836

QY 2226 GGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAG 2285

Db 835 GGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAAGAAAG 776

QY 2286 TTCGCCCATTTAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTCTGGAA 2345

Db 775 TTCGCCCATTTAGCCGAGAGATCACAATGAGCCCAAGCATATCAGAACATGTGTCTGGAA 716

QY 2346 TGTTTAAACCATGGTAGCATTGTGACATGGACGGCAAGTAGCTGTAACCGAAAGTTTGAGC 2405

Db 715 TGTTTAAACCATGGTAGCATTGTGACATGGACGGCAAGTAGCTGTAACCGAAAGTTTGAGC 656

QY 2406 TTGATAGTGAACAAAGTTCGGTATGAACACAGAGTTTGCTCCATTCAACAGTGTGATGACCC 2465

Db 655 TTGATAGTGAACAAAGTTCGGTATGAACACAGAGTTTGCTCCATTCAACAGTGTGATGACCC 596

QY 2466 CGCCGCCAGTGCACACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTC 2525

Db 595 CGCCGCCAGTGCACACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTC 536

QY 2526 CTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCCAAAAATGA 2585

Db 535 CTCCTCAGTCTCCTGAACCTGTATGTGGCAGCTAGTAAGCACTTTCAACAGGCCAAAAATGA 476

QY 2586 TATTGGAAAAATATTCTTAACCCGGACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAAC 2645

Db 475 TATTGGAAAAATATTCTTAACCCGGACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAAC 416

QY 2646 CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCTC 2705

Db 415 CCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCCTC 356

QY 2706 CTGAATTTGATTTCTCTGCTCATAAATATTTTCTGTGTGAAACTTGTTTGAGAGAGAC 2765

Db 355 CTGAATTTGATTTCTCTGCTCATA--ATATTTCTGTGTGAAACTGTTGAGAGAGACTG 299

QY 2766 TGGGGAGG 2773

Db 298 GGAGGTGG 291

RESULT 4

US-09-814-353-15927/c

; Sequence 15927, Application US/09814353

; Publication No. US20030165831A1

; GENERAL INFORMATION:

; APPLICANT: Lee, John

; APPLICANT: Thompson, Pamela

; APPLICANT: Lillie, James

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER

; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US 60/191,031

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: US 60/207,124

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: US 60/211,940

; PRIOR FILING DATE: 2000-06-15

; PRIOR APPLICATION NUMBER: US 60/216,820

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: US 60/257,672

; PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15927

; LENGTH: 745

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 571

; OTHER INFORMATION: n = A,T,C or G

US-09-814-353-15927

Query Match 18.1%; Score 560.8; DB 10; Length 745;

Best Local Similarity 94.8%; Pred. No. 1.3e-142;

Matches 621; Conservative 0; Mismatches 8; Indels 26; Gaps 3;

QY 1399 GGCATCCAGGCCCCAGAAATGATACTACAAAAAGGAGATCATCCCAATTATGATGGGTTTGAA 1458

Db 739 GGCAATCCAGGCCCCAGAAATGATACTACAAAAAGGAGATCATCCCAATTATGATGGGTTTGAA 680

QY 1459 CCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATTAA 1518

Db 679 CCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGCAAAAATAATTAA 620

QY 1519 AGGGAAGAAATGGTGAACATAATTTTCAAGATTAAATAGATAGATAAANAATCTGTCTGTGAG 1578

Db 619 AGGGAAGAAATGGTGAACATAATTTTCAAGATTAAATAGATAGATAAANAATCTGTCTGTGAG 560

QY 1579 GTTGTGAATTTAAACAAATTTACATTGTATCCTGGA-TTTTTTCTGTGAATTTAGTGAACA 1637

Db 559 GTTGTGAATTTAAACAAATTTACATTGTATCCTGGATTTTTTCTGTGAATTTAGTGAACA 500

QY 1638 GTCACCATGTGTTCTTTCAAGATCTCTGTGTACAAACCACCTTTCCTGGTGGATAACAAAA 1697

Db 499 GTCACCATGTGTTCTTTCAAGATCTCTGTGTACAAACCACCTTTCCTGGTGGATAACAAAA 440

QY 1698 GGTCCTTTGGAACCTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCTTTTGTGAG 1757

Db 439 GGTCCTTTGGAACCTCATCTCATGCAAGACATGGTGAAGATGCACTTCGGTCTTTTGTGAG 380

QY 1758 ATCCTCCGAGTGCTTTCCCCCAAGTGTCTATATATAATACACAGGCTAAGGACTGT 1817

Db 379 TCCTCCG--GTGCTTTCCCCCAAGTGTACCTATATAATAATACACAGGCTAAGGACTGT 322

QY 1818 ATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTTCAGATCCATGGA 1877

Db 321 ATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTTCAGATCCATGGA 262

QY 1878 CATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATATTTCTTGAGGAATTTGCCACCTTG 1937

Db 261 CATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAATATTTCTTGAGGAATTTGCCACCTTG 202

QY 1938 CAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAGGTTGATGCAGCGCTTCACA 1997

Db 201 CAGGATG-----AGGCAGAGAGGTTGATGCAGCGCTTCACA 165

QY 1998 CCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGTACCTG 2052

Db 164 CCATGCTGTTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTTAGGTACCTG 110

RESULT 5

US-09-814-353-16155

; Sequence 16155, Application US/09814353

; Publication No. US20030165831A1

; GENERAL INFORMATION:

; APPLICANT: Lee, John

; APPLICANT: Thompson, Pamela

; APPLICANT: Lillie, James

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER

; FILE REFERENCE: MRI-006B

; CURRENT APPLICATION NUMBER: US/09/814,353

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US 60/191,031

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: US 60/207,124

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: US 60/211,940

; PRIOR FILING DATE: 2000-06-15

; PRIOR APPLICATION NUMBER: US 60/216,820

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/220,661

; PRIOR FILING DATE: 2000-07-25

; PRIOR APPLICATION NUMBER: US 60/257,672

; PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15927

; LENGTH: 745

```
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16155
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-16155

Query Match          16.5%; Score 509.4; DB 10; Length 610;
Best Local Similarity 99.8%; Pred. No. 1.4e-128;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1184 GTACTCGAAGTCGACAAAGGAGAAGAGATCCAGAAGTTGAACTAGAACACCAACAAT 1243
Db 100 GTACTCGAAGTCGACAAAGGAGAAGAGATCCAGAAGTTGAACTAGAACACCAACAAT 159

QY 1244 GTTTAGCAGTATTGACGAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCT 1303
Db 160 GTTTAGCAGTATTGACGAGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCT 219

QY 1304 TTACTAAGAAAGACGACGAGTGCTGTTGAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAG 1363
Db 220 TTACTAAGAAAGACGACGAGTGCTGTTGAGAAAGCTCAAAAATTGATGGTTCAAGCAGCAG 279

QY 1364 ATCTTCTTTCTGCCATTTCATAATTTCATTGTCATCGGCATCCAGGCCAGAAATGATACTA 1423
Db 280 ATCTTCTTTCTGCCATTTCATAATTTCATTGTCATCGGCATCCAGGCCAGAAATGATACTA 339

QY 1424 CAAAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTC 1483
Db 340 CAAAAGGAGATCATCCAATTATGATGGGTTTTGAACCCCTTGTGAAACCAGAGGCTACTTC 399

QY 1484 CACCTACTTCCCTCGATATGCAAAAATAAATTAAGGGGAAGAAATGGTGAACATAATTTTG 1543
Db 400 CACCTACTTCCCTCGATATGCAAAAATAAATTAAGGGGAAGAAATGGTGAACATAATTTTG 459

QY 1544 CAAGATTAATAGATAGATAAATAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATT 1603
Db 460 CAAGATTAATAGATAGATAAATAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATT 519

QY 1604 GTATCCTGGATTTTTCTGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTC 1663
Db 520 GTATCCTGGATTTTTCTGTGAATTTAGTGAAACAGTCACCATGTGTTCTTTCAAGATCTC 579

QY 1664 TGTACAAACCACCTTTCCCTGGTGGATAACAA 1694
Db 580 TGTACAAACCACCTTTCCCTGGTGGATAACAA 610
```

RESULT 6  
US-09-814-353-3218/c  
; Sequence 3218, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21

```
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3218
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 561,590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-3218
```

```
Query Match          16.0%; Score 496.2; DB 10; Length 602;
Best Local Similarity 93.9%; Pred. No. 5.8e-125;
Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;

QY 1450 GGTTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC-AAA 1508
Db 602 GGTTTTGAACCCNTTGTGAACCCAGAGGCTACTTCCACCTACNTTCCCTCGATATGCAAA 543

QY 1509 AATAATTAAAGGGAAGAAATGGTGAACATAATTTGCAAGATTAAATAGATAGAATAAAAC 1568
Db 542 AATAATTAAAGGGAAGAAATGGTGAACATAATTTGCAAGATTAAATAGATAGAATAAAAC 483

QY 1569 TGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCTGGA-TTTTTCTGTGAAT 1627
Db 482 TGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCTGGATTTTTTCTGTGAAT 423

QY 1628 TTAGTGAACAGTCACCAATGTTCTTTCAAGATCTCTGTACAAACCACCTTTCCTGGTGG 1687
Db 422 TTAGTGAACAGTCACCAACGTTCTTTCAAGATCTCTGTACAAACCACCTTTCCTGGTGG 363

QY 1688 ATAACAAAAAGGTCTTTGGAACTCATCTCATGAAGACATGGTGAAGATGCACCTTCGGT 1747
Db 362 ATAACAAAAAGGTCTTTGGAACTCATCTCATGAAGACATGGTGAAGATGCACCTTCGGT 303

QY 1748 CTTTTTGTGAGATCCTCCGAGTGCTTTCCCCAAAGTGCTACCTATATAATAATCACCAGGC 1807
Db 302 CTTTTTGTGAG--TCCTCCGGTGCTTTCCCCAAAGTGCTACCTATATAATAATCACCAGGC 245

QY 1808 TAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGGCCATTCTGTAGTCTTATTCA 1867
Db 244 TAAGGACTGTATCGACTCCTTTGTTTACTCACTGTGTTGGCCATTCTGTAGTCTTATTCA 185

QY 1868 GATCCATGGACATAACAGGGCTCGACAGAGATAAGCTTGGTCATATTCTTGAGGAATT 1927
Db 184 GATCCATGGACATAACAGGGCTCGACAGAGATAAGCTTGGTCATATTCTTGAGGAATT 125

QY 1928 TGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGAAGGTTGATGCA 1987
Db 124 TGCCACCTTGCAGGATG-----AGGCAGAGAAGGTTGATGCA 88

QY 1988 GCGCTTCACACCATGCTGTGTAACAGGAACCCCAAGGCAACATTTGGCCTGTTAGGT 2047
Db 87 GCGCTTCACACCATGCTGTGTAACAGGAACCCCAAGGCAACATTTGGCCTGTTAGGT 28

QY 2048 ACCTG 2052
Db 27 ACCTG 23
```



RESULT 7  
US-09-814-353-9543/c  
; Sequence 9543, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9543  
; LENGTH: 602  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 561..590  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-9543

Query Match 16.0%; Score 496.2; DB 10; Length 602;  
Best Local Similarity 93.9%; Pred. No. 5.8e-125;  
Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;  
QY 1450 GGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACCTTCCCTCGATATGC-AAA 1508  
Db |||||||  
602 GGTTTGAACCCNTTGTGAACCCAGAGGCTACTTCCACCTACNTTCCCTCGATATGCAAAA 543  
QY 1509 AATAATTAAAGGGAAGAAATGGTGAACATATTTTGCAAGATTAAATAGATAGAATAAAAAC 1568  
Db |||||||  
542 AATAATTAAAGGGAAGAAATGGTGAACATATTTTGCAAGATTAAATAGATAGAATAAAACC 483  
QY 1569 TGTCTGTGAGGTGTGAATTTAACAAATTTACATTTGATCTCTGGA-TTTTTCTGTGAAT 1627  
Db |||||||  
482 TGTCTGTGAGGTGTGAATTTAACAAATTTACATTTGATCTCTGGAATTTTCTGTGAAT 423  
QY 1628 TTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGTGG 1687  
Db |||||||  
422 TTAGTGAACAGTCACCATGTGTTCTTTCAAGATCTCTGTTACAAACCACTTTCCTGGTGG 363  
QY 1688 ATAACAAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCGGT 1747  
Db |||||||  
362 ATAACAAAAAGGCTTTGGAACTCATCTCATGCAAGACATGGTGAAGATGCACCTTCGGT 303  
QY 1748 CTTTGTGCAGATCCTCCGAGTGCTTTCCCCCAAGTGCTACCTATATAATAATCACCAGGC 1807  
Db |||||||  
302 CTTTGTGCAG--TCCTCCGGTGCTTTCCCCCAAGTGCTACCTATATAATAATCACCAGGC 245  
QY 1808 TAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTCTCTAGTCTTATTCA 1867  
Db |||||||  
244 TAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCCGCCATTCTCTAGTCTTATTCA 185  
QY 1868 GATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAATT 1927  
Db |||||||  
184 GATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTTCTTGAGGAATT 125

QY 1928 TGCCACCTTGCCAGGTAGTTTATGACATTTTATTAATAGGCAGAGAAGTTTGATGCA 1987  
Db |||||||  
124 TGCCACCTTGCCAGGTAG--AGGCAGAGAAGTTTGATGCA 88  
QY 1988 GCGCTTCACACCATGCTGTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTAGGT 2047  
Db |||||||  
87 GCGCTTCACACCATGCTGTGAAACAGGAACCCCAAGGCAACATTTGGCCTGTTAGGT 28  
QY 2048 ACCTG 2052  
Db |||||  
27 ACCTG 23  
RESULT 8  
US-09-918-995-20914  
; Sequence 20914, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 20914  
; LENGTH: 482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(482)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-20914

Query Match 14.6%; Score 451.2; DB 10; Length 482;  
Best Local Similarity 99.3%; Pred. No. 1.1e-112;  
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 491 CGTCGTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGGGTGACCACGGGAGAAGTAG 550  
Db |||||||  
26 CGAGCTTATTTCCGTGGTCCGGACAGTGCCTGGCGCGGGTGACCACGGGAGAAGTAG 85  
QY 551 GCATAATGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAGTATGC 610  
Db |||||||  
86 GCATAATGTTATGAAAGCTTCTGTAGATGATGACGATTTCAGGATGGGAGCTCAGTATGC 145  
QY 611 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTACCAAGATTTTGAAGAAG 670  
Db |||||||  
146 CAGAAAAATGGAGAAAAAGCAATACAACTGGGTGGACATTACCAAGATTTTGAAGAAG 205  
QY 671 CTTGTCGAGAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGTCTTTTGAAG 730  
Db |||||||  
206 CTTGTCGAGAATTAAAGTTGGGAGAACTACTTTCATGATAAGCTATTTGGTCTTTTGAAG 265  
QY 731 CCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCAAG 790  
Db |||||||  
266 CCATGTCGTCTATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAACCAAG 325  
QY 791 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAATAAATAAAG 850  
Db |||||||  
326 TTAATCGAAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAATAAATAAAG 385  
QY 851 ATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTTGCTGTTTGAACCGT 910  
Db |||||||  
386 ATCTCACCTTGCCTGAACTGATAGGGATTATGGATACATGTTTTGCTGTTTGAACCGT 445  
QY 911 GGTTAGAGGCCATTCACTGGCACAGACAGTATTTA 946  
|||

Db 446 GGTTAGAAGGCCAATTCACCTGGCACAGACAGATATTTA 481

RESULT 9

US-09-814-353-3455  
; Sequence 3455, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3455  
; LENGTH: 434  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-814-353-3455

Query Match 14.0%; Score 434; DB 10; Length 434;  
Best Local Similarity 100.0%; Pred. No. 5.3e-108;  
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1186	ACTCGAAGTCGACAAGGAGAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT	1245
Db	1	ACTCGAAGTCGACAAGGAGAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT	60
Qy	1246	TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGCAGTGCTTATAGCCTTT	1305
Db	61	TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGCAGTGCTTATAGCCTTT	120
Qy	1306	ACTAAGAAAGAGACCAGTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAT	1365
Db	121	ACTAAGAAAGAGACCAGTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAT	180
Qy	1366	CTTCTTTCTGCCATTCAATAATTCATTCATGGCATCCAGGCCCAGAAATGATACTACA	1425
Db	181	CTTCTTTCTGCCATTCAATAATTCATTCATGGCATCCAGGCCCAGAAATGATACTACA	240
Qy	1426	AAAGGAGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	1485
Db	241	AAAGGAGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	300
Qy	1486	CCTACCTTCCCTCGATATGCAAAAATAATTAAGGGGAAGAAATGGTGAACCTATTTGCA	1545
Db	301	CCTACCTTCCCTCGATATGCAAAAATAATTAAGGGGAAGAAATGGTGAACCTATTTGCA	360
Qy	1546	AGATTAAATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTTACATTGT	1605
Db	361	AGATTAAATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTTACATTGT	420
Qy	1606	ATCCTGGATTTTTT 1619	
Db	421	ATCCTGGATTTTTT 434	

RESULT 11

RESULT 10

US-09-814-353-9771  
; Sequence 9771, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 9771  
; LENGTH: 434  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-814-353-9771

Query Match 14.0%; Score 434; DB 10; Length 434;  
Best Local Similarity 100.0%; Pred. No. 5.3e-108;  
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1186	ACTCGAAGTCGACAAGGAGAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT	1245
Db	1	ACTCGAAGTCGACAAGGAGAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT	60
Qy	1246	TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGCAGTGCTTATAGCCTTT	1305
Db	61	TTAGCAGTATTCAGCAGAGTGAAATTTACTCGTGTGTTACTGCAGTGCTTATAGCCTTT	120
Qy	1306	ACTAAGAAAGAGACCAGTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAT	1365
Db	121	ACTAAGAAAGAGACCAGTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAT	180
Qy	1366	CTTCTTTCTGCCATTCAATAATTCATTCATGGCATCCAGGCCCAGAAATGATACTACA	1425
Db	181	CTTCTTTCTGCCATTCAATAATTCATTCATGGCATCCAGGCCCAGAAATGATACTACA	240
Qy	1426	AAAGGAGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	1485
Db	241	AAAGGAGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCA	300
Qy	1486	CCTACCTTCCCTCGATATGCAAAAATAATTAAGGGGAAGAAATGGTGAACCTATTTGCA	1545
Db	301	CCTACCTTCCCTCGATATGCAAAAATAATTAAGGGGAAGAAATGGTGAACCTATTTGCA	360
Qy	1546	AGATTAAATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTTACATTGT	1605
Db	361	AGATTAAATAGATAGAATAAAAACCTGCTGTGAGGTTGTGAATTTAACAAATTTACATTGT	420
Qy	1606	ATCCTGGATTTTTT 1619	
Db	421	ATCCTGGATTTTTT 434	

US-09-983-965-210  
; Sequence 210, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 210  
; LENGTH: 398  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 25-BOVMS1-018-Q1-E1-G1  
US-09-983-965-210

Query Match 11.3%; Score 350.8; DB 9; Length 398;  
Best Local Similarity 94.3%; Pred. No. 3.4e-85;  
Matches 364; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
  
QY 2298 GCCGAGAGATCACAATGAGCCAGCATATCAGAACATGTGTGCTGGAATGTTTAAACCA 2357  
|||||  
Db 13 GCCGAGAGATCACAATGAGCCAGCATATCAGAACATGTGTGCTGGAATGTTTAAACCA 72  
  
QY 2358 TGGTAGATTGACATGGACGGCAAGTACGTAAACCGAAGTTTGAGCTTGATAGTGAC 2417  
|||||  
Db 73 TGGGAGCTTTGACATGGATGGCAAGTACGAAACCCCAAGTTTGAGCTTGATAGTGAC 132  
  
QY 2418 AAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGCCAGTGC 2477  
|||||  
Db 133 AAGTTCGATATGACACAGATTTGCTCCATTCAACAGTGTATGACACCCACCAGTGC 192  
  
QY 2478 ACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCAGTCTC 2537  
|||||  
Db 193 ACTATCTGCAGTTCAGGAAATGTCTGACCTCAATAAATATAGCCCTCCTCAGTCTC 252  
  
QY 2538 CTGAACTGTATGTGGCAGCTAGTAAGCACCTTCAACAGGCAAAATGATATTGGAAATA 2597  
|||||  
Db 253 CAGAACTGTATGTGGCAGCTAGTAAGCACCTTTCAGCAGGCAAAATGATACTGGAAATA 312  
  
QY 2598 TTCTTAACCCGACCATGAGGTTAATAGAAATTTTAAAGTTGCCAAACCCCACTTTGIGG 2657  
|||||  
Db 313 TCCCAAAACCCAGACCATGAGGTCAATAGAAATTTTAAAGTTGCCAAACCCCACTTTGIGG 372  
  
QY 2658 TTATGAAGTTATTGGCAGGAGGACAC 2683  
|||||  
Db 373 TTATGAAGTTATTGGCAGGAGGACAC 398

RESULT 12  
US-10-085-783A-43407  
; Sequence 43407, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017

; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43407  
; LENGTH: 345  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (16)..(16)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-085-783A-43407  
  
Query Match 9.8%; Score 302.6; DB 13; Length 345;  
Best Local Similarity 97.3%; Pred. No. 5.2e-72;  
Matches 329; Conservative 0; Mismatches 5; Indels 4; Gaps 2;  
  
QY 2747 AAACCTGTTTGAGAGAGACTCGGGAGGTGGCCATAAAGGGCGAGAGTCTTCTTTTCAGACC 2806  
|||||  
Db 1 AACCTGTTTGAGAGAGACTCGGAAGGTGGCCCTAAAGGGCGAGAGTCTTCTTTTCAGACC 60  
  
QY 2807 CAACTCTTAGAGGGCCACATCACACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTGG 2866  
|||||  
Db 61 CAACTCTTAGAGGGCCACATCACACAGGCTCCACATCACGGGAAGTGAGATGGATTTCTGG 120  
  
QY 2867 GTAACAACTCATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAA 2926  
|||||  
Db 121 GTAACAACTCATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAA 180  
  
QY 2927 CTGCTGTTTAAAGTGTTTATTTATGTTCATGGAAGAACTGGTCTTATTGAATGCATT 2986  
|||||  
Db 181 CTGCTGTTTAAAGTGTTTATTTATGTTCATGGAAGAACTGGTCTTATTGAATGCATT 240  
  
QY 2987 GATGAACGTTATAT--GGTTTATTACAGATTTAATCACAATCATTTTATGAATGAT 3044  
|||||  
Db 241 GATGAACGTTATATTTGGTTTATTACAGATTTAATCACAATCATTTTATGAATGAT 300  
  
QY 3045 TGAGTGAAAATAGTGTTTTATAAA--GGTTAATAAATTT 3080  
|||||  
Db 301 TGAGTGAAAATAGTGTTTTATAAAAGGGTTAATAAATTT 338  
  
RESULT 13  
US-10-242-535A-43407  
; Sequence 43407, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43407  
; LENGTH: 345  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (16)..(16)  
; OTHER INFORMATION: n is a, c, g, or t



```
US-10-242-535A-43407
Query Match          9.8%; Score 302.6; DB 16; Length 345;
Best Local Similarity 97.3%; Pred. No. 5.2e-72;
Matches 329; Conservative 0; Mismatches 5; Indels 4; Gaps 2;

QY 2747 AAACCTGTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACC 2806
Db 1 AACCTTGTTTGAGAGNAGACTGGGAAGGTGGCCCTAAAGGGGCAGAGTCTTCTTTCAGACC 60

QY 2807 CAACTCTTAGAGGGCACATCACCCAGGCTCCACATCACGGGAAGTGAGATGGATTCTTGG 2866
Db 61 CAACTCTTAGAGGGCACATCACCCAGGCTCCACATCACGGGAAGTGAGATGGATTCTTGG 120

QY 2867 GTAACAACTCATATAAGGAATACCTTTTAGTTTGACAGCCCTTATATGACATGAATGAAA 2926
Db 121 GTAACAACTCATATAAGGAATACCTTTTAGTTTGACAGCCCTTATATGACATGAATGAAA 180

QY 2927 CTGCTGTTTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATT 2986
Db 181 CTGCTGTTTTAAAGTGGTTTATTATGTTCCATGGAAGAAACTGGTCTTATTGAATGCATT 240

QY 2987 GATGAACGTTATAT--GGTTTTATTACAGATTTAATCACAAATCATTTTTTATGAATGAT 3044
Db 241 GATGAACGTTATATGGTTTTTTATTACAGATTTAATCACAAATCATTTTTTATGAATGAT 300

QY 3045 TGAGTGAAATAAGTGTTTTATAAA--GGTTAATAAATTT 3080
Db 301 TGAGTGAAATAAGTGTTTTATAAAAGGGTTAATAAATTT 338
```

```
RESULT 14
US-09-783-590-5693
; Sequence 5693, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5693
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (353)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (383)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (403)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (425)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (440)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (449)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (455)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (466)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (479)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5693

Query Match          8.7%; Score 268.6; DB 9; Length 501;
Best Local Similarity 93.2%; Pred. No. 1.5e-62;
Matches 313; Conservative 0; Mismatches 17; Indels 6; Gaps 3;

QY 2223 AAAGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAGA 2282
Db 7 AAAGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAAAACAAGAAAAAGA 66

QY 2283 AAGTTGCCCATTTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAACATGTGTGCTG 2342
Db 67 AAGTTGCCCATTTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAACATGTGTGCTG 126

QY 2343 GAATGTTTAAACCATTGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTTTG 2402
Db 127 GAATGTTTAAACCATTGGTAGCATTTGACATGGACGGCAAGTACGTAAACCGAAGTTTG 186

QY 2403 AGCTT-GATAGTGAACAAGTTCCGTATGAACACACAGGTTTGCTCCATTCAACAGTGTGATG 2461
Db 187 AGCTTGGTTAGTGAACAAGTTCCGTATGAACACACAGGTTTNCCTCCATTCAACAGTGTGATG 246

QY 2462 ACCCGCGCCGAGTGCACTACTTACAGTT--CAAGGAAATGTTCTGACCTCAATAAATA 2519
Db 247 ACCCGCGCCGAGTGCACTACTTACAGTTTCAAGGAAATGTTTNTGACCTCAATAAATTTA 306

QY 2520 GCCCTCCTCC--TCAGTCTCCTGAACTGTATGTGG 2552
Db 307 GCCCTCCTNCCTTTCAGTTTTCCTGAACTGTATGTGG 342
```

```
RESULT 15
US-10-027-632-323193/c
; Sequence 323193, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
```

```
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 323193
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-323193
```

```
Query Match          5.9%; Score 182; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 8.8e-39;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1431 AGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTAC 1490
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 401 AGATCATCCAATTATGATGGGTTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTAC 342

Qy 1491 CTTCCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGCAAGATT 1550
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 341 CTTCCCTCGATATGCAAAAATAATTAAGGGAAGAAATGGTGAACCTATTTTGCAAGATT 282

Qy 1551 AATAGATAGAATAAAAACTGCTGTGAGGTTGTGAATTTTAACAATTTACATTGTATCCT 1610
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 281 AATAGATAGAATAAAAACTGCTGTGAGGTTGTGAATTTTAACAATTTACATTGTATCCT 222

Qy 1611 GG 1612
    ||
Db 221 GG 220
```

Search completed: August 11, 2004, 13:00:44  
Job time : 1439.92 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 17:10:22 ; Search time 225.025 Seconds  
(without alignments)  
7635.278 Million cell updates/sec

Title: US-10-001-857-42  
Perfect score: 3096  
Sequence: 1 ttcttcacgaaactcccagg.....attttttgacaaaaaaaaa 3096

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	%		Query Match	Length	DB ID	Description
	Score	Match				
C 1	69.4	2.2	7218	1	US-08-232-463-14	Sequence 14, Appl
2	56.4	1.8	320	3	US-09-165-264-11	Sequence 11, Appl
3	55.2	1.8	320	3	US-09-165-264-14	Sequence 14, Appl
4	54.8	1.8	320	3	US-09-165-264-7	Sequence 7, Appl
5	54.8	1.8	320	3	US-09-165-264-13	Sequence 13, Appl
6	54.2	1.8	318	3	US-09-165-264-12	Sequence 12, Appl
7	54.2	1.8	319	3	US-09-165-264-8	Sequence 8, Appl
8	54.2	1.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl
9	54.2	1.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 10	53.2	1.7	2561	4	US-09-616-289-48	Sequence 48, Appl
C 11	51	1.6	152331	3	US-09-128-155-16	Sequence 16, Appl
C 12	50	1.6	1614	4	US-09-616-289-45	Sequence 45, Appl
C 13	50	1.6	12425	4	US-09-616-289-50	Sequence 50, Appl
C 14	48.4	1.6	23673	4	US-09-773-816-1	Sequence 1, Appl
C 15	47.4	1.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 16	47.4	1.5	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 17	46.4	1.5	483	4	US-09-543-681A-4138	Sequence 4138, Ap
C 18	46.4	1.5	516	4	US-09-543-681A-4140	Sequence 4140, Ap
C 19	46.4	1.5	519	4	US-09-543-681A-4145	Sequence 4145, Ap
C 20	46.4	1.5	608	4	US-09-543-681A-4115	Sequence 4115, Ap
21	46	1.5	4638	4	US-09-023-655-1215	Sequence 1215, Ap
22	45.6	1.5	8961	4	US-10-204-708-80	Sequence 80, Appl
23	45.2	1.5	1674	4	US-09-894-844-14	Sequence 14, Appl
24	45.2	1.5	1674	4	US-09-894-844-15	Sequence 15, Appl
25	45	1.5	1561	4	US-08-765-907A-11	Sequence 11, Appl
26	44.8	1.4	2130	4	US-09-909-962A-6	Sequence 6, Appl
27	44.8	1.4	2130	4	US-09-909-962A-7	Sequence 7, Appl

C 28	44	1.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
29	44	1.4	3572	4	US-09-575-574-3	Sequence 3, Appl
C 30	44	1.4	12001	1	US-08-458-568A-11	Sequence 11, Appl
31	43.8	1.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
32	43.6	1.4	239	4	US-09-056-556-181	Sequence 181, App
33	43.6	1.4	239	4	US-09-072-596-176	Sequence 176, App
34	43.6	1.4	239	4	US-09-072-967-181	Sequence 181, App
C 35	43.6	1.4	2335	3	US-09-387-574-9	Sequence 9, Appl
C 36	43.6	1.4	2335	4	US-09-668-096-9	Sequence 9, Appl
C 37	43.4	1.4	1070	4	US-09-470-443-7	Sequence 7, Appl
C 38	43.4	1.4	1294	3	US-09-312-038-4	Sequence 4, Appl
C 39	43.4	1.4	1294	4	US-09-850-964-4	Sequence 4, Appl
C 40	43.4	1.4	2289	3	US-09-312-038-3	Sequence 3, Appl
C 41	43.4	1.4	2289	4	US-09-850-964-3	Sequence 3, Appl
C 42	43.4	1.4	5482	4	US-09-470-443-3	Sequence 3, Appl
43	43.2	1.4	1215	4	US-09-091-952A-29	Sequence 29, Appl
44	42.8	1.4	339	4	US-09-543-681A-4147	Sequence 4147, Ap
45	42.4	1.4	289	3	US-09-007-005-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 2.2%; Score 69.4; DB 1; Length 7218;



[illegible][illegible][illegible]

```

RESULT 4
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```



```
QY 423 ATACGCATCGTGCACGCTGCCGCTCGGCTGGGCTGAGAGGGAGGGGGCGCGCGGC 482
      |||||
Db 176 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 235
      |||||
QY 483 CGAGGCGGCGTCGTTATTTCCTGTCGGACAGTGCGTGGCGGCGCGGGTGACACGGG 542
      |||||
Db 236 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 295
      |||||
QY 543 AGAAGTAGGCA 553
      |||||
Db 296 GGGGGTCTGCA 306
      |||||
```

```
RESULT 8
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 1.8%; Score 54.2; DB 3; Length 4403765;
Best Local Similarity 52.4%; Pred. No. 0.016;
Matches 119; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 318 GGCGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGGTGTGGCCGGGAGCGCGAA 377
      |||||
Db 3937877 GCGGCGATGGCGGCAACGGCGCAACGGCGGTACCGGCGGCGAGGGCGGAGTCCGC 3937936

QY 378 GTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGCTCCCGGCATACGCATGCGTGA 437
      |||||
Db 3937937 GCGCGCGGGCACCGGCGGTGCGGCCGCAATGGCGGCGATGGCGGCACCGCGGCACC 3937996

QY 438 CGTCCCGTCCGGCTGGGCTGAGAGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 497
      |||||
Db 3937997 GCGCGGGGGACGGGGGCGCGCGGCACCGCGGCACCGCGGCGACCGCGGCCTCGGC 3938056

QY 498 ATTTCCGTGTCGGACAGTGCGTGGCGGCGCGGGTGACCAACGGGAG 544
      |||||
Db 3938057 GACCCCGGTCGCGCGATCCGCGGCGACGCGGCACCGCGGCAG 3938103
```

```
RESULT 9
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
```

```
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1
```

```
Query Match 1.8%; Score 54.2; DB 3; Length 4411529;
Best Local Similarity 52.4%; Pred. No. 0.016;
Matches 119; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 318 GGCGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGGTGTGGCCGGGAGCGCGAA 377
      |||||
Db 3944330 GCGGCGATGGCGGCAACGGCGCAACGGCGGTACCGGCGGCGAGGGCGGAGTCCGC 3944389

QY 378 GTCCCGGGAGTAAGGGAGAGGGGCGGGGTGCGCGCTCCCGGCATACGCATGCGTGA 437
      |||||
Db 3944390 GCGCGCGCGGCACCGGCGGTGCGGCCGCAATGGCGGCGATGGCGGCACCGCGGCACC 3944449

QY 438 CGTCCCGTCCGGCTGGGCTGAGAGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 497
      |||||
Db 3944450 GCGCGGGGGACGGGGGCGCGCGGCACCGGCGGCACCGCGGCGCCTCGGC 3944509

QY 498 ATTTCCGTGTCGGACAGTGCGTGGCGGCGCGGGTGACCAACGGGAG 544
      |||||
Db 3944510 GACCCCGGTCGCGCGATCCGCGGCGACGCGGCACCGCGGCAG 3944556
```

```
RESULT 10
US-09-616-289-48/c
; Sequence 48, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 2561
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (246)...(1895)
US-09-616-289-48
```

```
Query Match 1.7%; Score 53.2; DB 4; Length 2561;
Best Local Similarity 52.8%; Pred. No. 0.00023;
Matches 115; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 335 CGGAGGCGGTGGCGAGGGGGGTGTGGCCGGGAGCGCGGAAGTCCCGGGAGTAAGGG 394
      |||||
Db 861 CGCGGGCGGGGGGCGCGCGCGGGCCCGCGGGGGCGCGCCCGGGGGAGCGGCGG 802
```





US-09-616-289-50

Query Match 1.6%; Score 50; DB 4; Length 12425;  
Best Local Similarity 53.9%; Pred. No. 0.0051;  
Matches 124; Conservative 0; Mismatches 105; Indels 1; Gaps 1;  
  
QY 315 GCGGCGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGTGTGGCCGGGAGCGC 374  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 3460 GCGGGCTCCGGCGCGACGGCGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCG 3402  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 375 GAAGTCCCGGAGTAAGGAGAGGGGGCGGGTCCCGGGCATACGATGGT 434  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 3401 CACCGCGGGGAGCGGCTGGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCG 3342  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 435 GCACGTGCGCGTGGGCTGAGAGGGGAGGGGCGGCGCGCGCGCGCGCGCGTC 494  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 3341 GCGGGCGCGGCTGCGCGGGCCAGCGAGGGGGCGCTGTGGCGGCGCGCGCGCG 3282  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 495 GTTATTTCCGTGTCGGACAGTGCCTGGCGCGCGGGTGACACGGGAG 544  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 3281 GCGGGCGCGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGGCG 3232  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14  
US-09-773-816-1/c  
; Sequence 1, Application US/09773816  
; Patent No. 6340774  
; GENERAL INFORMATION:  
; APPLICANT: Stanford University  
; APPLICANT: Khosla, Chaitan  
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR  
; TITLE OF INVENTION: ANTAGONISTS  
; FILE REFERENCE: 28600-20210.00  
; CURRENT APPLICATION NUMBER: US/09/773,816  
; CURRENT FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: US 60/243,458  
; PRIOR FILING DATE: 2000-10-25  
; PRIOR APPLICATION NUMBER: US 60/179,305  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 23673  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(23623)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-773-816-1

Query Match 1.6%; Score 48.4; DB 4; Length 23673;  
Best Local Similarity 46.7%; Pred. No. 0.022;  
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
  
QY 182 CGCAGACGGGCGGGCTCTGAGACTCCGGCTCCGCCCTCTTTCCGGGAACCGCCCACTAC 241  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 5007 CGCATGCGCGCGCGCTCGACCAGTTCGAGGTCGAGGGCCAAACGGGTGCGCACCATC 4948  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 242 CCAGGACTCCGACAGAGGTTGAAAAAAGATAACTTCGGTCTCGGATCGTCTAATCT 301  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 4947 GGGTTCCTGCGCAGGTGTTTCGACACGACAAGTTCGCCGCCGAGCGGACGACACCTCC 4888  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 302 CGCGAGAGAGAAGGCGCGCCCATCGGCCGAACGGAGCGGTGGCGAGGGGGGTGT 361  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 4887 CTCGTGACGAGATGCTCGCCCGAGGCCCGGAACGCCGAGTGCCTGTTC 4828  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 362 GCGCGGGAGCCGGAAGTCCCGGGAGTAAGGAGAGGGGGCGGGTCCGCGTCCCGGG 421  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 4827 GCGTGAAGCGGTGCGCGGTGCCCGGCGGTCTCCAGGCAACGCCGACCCCTCCGCC 4768  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 422 CATACGATGCGTGACGCTGCCCGTGGGCTGAGAGGGAGGGGCGGCGCGG 481  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 4767 GGCCCGCGAACGACGACGCCGCGGGCCGGGGGTTCGCGTACCGGAGCGGGG 4708  
  
QY 482 CCGAGGCGGCGTCTGTTATTTCCTGGTCCG 511  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 4707 CCGGCCCCGAACGGGATGTCCCGGACCCG 4678  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
  
RESULT 15  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
  
Query Match 1.5%; Score 47.4; DB 3; Length 4403765;  
Best Local Similarity 50.7%; Pred. No. 1.3;  
Matches 114; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
  
QY 317 CGGCCGCCATCGGCCGAACGGAGCGGTGGCGAGGGGGTGTGGCCGGGAGCGCGA 376  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2800482 CGGTGCGAGTTGGTGGCCGCCCGGTGACGGGGGGCGCGGTACGGTGGCGACCCGG 2800423  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 377 AGTCCCCGGGAGTAAGGAGAGGGGGCGGGTCCGGCTCCCGGCATACGATGCGTGC 436  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 2800422 CCTGGTGGGCTGGCGGAGCCGGGGGGCGTCCGGCGGCGCGGTGCTCGCGGGGGC 2800363  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 437 ACGTCCCGGTCCGGTGGGTGAGAGGGGAGGGGCGCGCGCGCGCGCGTCCGT 496  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 2800362 CAATGCTTGGCGGCCGCAACGACGGGCCGGTCAAGCGCGGCAACGGCGCAAGGTGG 2800303  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 497 TATTTCCGTGTCGGACAGTGCCTGGCGCGCGGTGACCAACCG 541  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 2800302 CAATGGCGCCACGACCGGTCCCGCGCGGTGCGCGGTATGGCGGTACCG 2800258  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: August 10, 2004, 21:58:26  
Job time : 236.025 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:49:02 ; Search time 20 Seconds  
(without alignments)  
3525.419 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MWKASVDDDDSGWELSMPE.....KVPPEFDFS AHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	584	15.2	695	B84495	hypothetical prote
2	548	14.2	814	T32223	hypothetical prote
3	270.5	7.0	733	S31288	MAK10 protein - ye
4	252.5	6.6	708	T39742	hypothetical prote
5	134	3.5	2429	1 SJHUA	spectrin alpha cha
6	132.5	3.4	1152	2 S20106	hypothetical prote
7	127	3.3	969	2 E71810	type III restricti
8	127	3.3	1089	2 S48244	NMD2 protein - yea
9	123.5	3.2	778	2 T16111	hypothetical prote
10	122.5	3.2	1607	2 T21982	hypothetical prote
11	122.5	3.2	2253	2 T30336	nuclear/mitotic ap
12	121.5	3.2	1603	1 VJKW5	vitellogenin vit-5
13	119	3.1	1196	2 T13057	KIAA0729 protein -
14	118.5	3.1	1029	2 T02576	hypothetical prote
15	117.5	3.1	3660	1 S02041	dystrophin, muscle
16	116	3.0	736	2 T00023	transcription fact
17	116	3.0	3147	2 T21328	hypothetical prote
18	114.5	3.0	1603	2 F89497	protein vit-5 [imp
19	114	3.0	1023	2 JC4013	major acidic nucle
20	114	3.0	1558	2 B71603	RESA-H3 antigen PF
21	114	3.0	1679	2 S48385	hypothetical prote
22	113.5	2.9	719	2 T05384	beta-adaptin homol
23	113.5	2.9	833	2 T43446	hypothetical prote
24	113.5	2.9	1957	2 T38077	hypothetical coile
25	113	2.9	2787	2 S45416	TEL1 protein - yea
26	113	2.9	3744	2 S46715	hypothetical prote
27	112.5	2.9	1642	2 T08880	NMDA receptor-bind
28	112	2.9	1130	2 T23104	hypothetical prote
29	112	2.9	1133	2 T23103	hypothetical prote

30	111.5	2.9	1505	2	JC4851	hypoxia-inducible
31	111	2.9	607	2	S48326	hypothetical prote
32	111	2.9	1067	2	S33417	kinesin-like prote
33	110.5	2.9	601	2	T49752	hypothetical prote
34	109.5	2.8	564	2	A60115	M protein precurs
35	109.5	2.8	1066	1	A48669	kinesin-related pr
36	109.5	2.8	1620	2	S61535	nucleotide-binding
37	109.5	2.8	2427	2	T16613	hypothetical prote
38	109	2.8	900	2	T50773	translation initia
39	109	2.8	1275	2	S65824	reverse transcript
40	109	2.8	1576	2	T21172	hypothetical prote
41	108.5	2.8	1442	2	C82898	DNA polymerase III
42	108.5	2.8	2672	2	A48126	translation activa
43	108	2.8	837	2	D69171	intracellular prot
44	108	2.8	1388	2	S70633	serine/threonine-s
45	108	2.8	1953	2	S63244	BN11 protein - yea

ALIGNMENTS

RESULT 1

B84495  
hypothetical protein At2g11000 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84495  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84495  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-695 <STO>  
A;Cross-references: GB:AE002093; NID:g4662644; PIDN:AAD26914.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g11000  
A;Map position: 2

Query Match	15.2%	Score 584;	DB 2;	Length 695;
Best Local Similarity	23.5%;	Pred. No. 1.1e-35;		
Matches	184;	Conservative	153;	Mismatches 291; Indels 156; Gaps 25;
QY	3	MKASVDDDDSGWELSMPEKMS-----NTNWVDITQDFEEACRELKLGELLHKLFLGL 56		
Db	1	MQSVREDEDDSSPIHHDSTSSITPSGDNNNSVWADVSPLLSACSDLQEGELINGDNFNL 60		
QY	57	FEAMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKI-----KDLTLPELIGIMDT 111		
Db	61	FAMSALEIMDPKMDSGMVST-----FYSIDEAIESGFAPVPISDSTVNVQSIIDIMDH 115		
QY	112	CFCCCLITWLEGHSLAQTVFTCLYIHNPDFIEDPA-MKAFALGILKICDIAREKVNKAAYF 170		
Db	116	LLACEATWHMGHSLAQTVFSCIYVLRPERTSSQALLHSYCRVIRATCRAVSVVSDARTN 175		
QY	171	EEEDFQSMTYGFKMANSVTDLRVTGMLKDVEDDMQRRV---KSTRSRQGEERDPEVE--- 224		
Db	176	EEEDLFTMTYGLPFSGE-EDAKGLLLINAVEETICQLRACKATRRRMLE--DAELEPLQ 232		
QY	225	----LEHQOCLAVFSRVKFRVLLTVLIATFKKETSAAVEAQKLMVQAADLLSAIHNS--- 278		
Db	233	SNPHLEESFCKSLLCRIRFRKHFHLHALNCMRRPQQRGLGLARKHIGYCISELDSVLDSAE 292		
QY	279	-----LHHGIQ--AQNDTTKGHPIMMGFEPLVNQRLPPTFPFYAKI IKREEMVNYFA 330		
Db	293	FLRLDIFENGVNEIEESTTAGRSPI--GFDPTLNKRLSAPTPPRAIKLLSWKKAIDYV 350		
QY	331	RLIDRIKTVCEVWNLTNLHLCILDFCFESEQSPCVLSRSLLOTFTFLVDNKKVFGTHLMQD 390		
Db	351	KLHLNLDKICAFSLEPDLAVLEFVFIQFKSRPDLVARAHLQ-LLLVQDGKLYGR----- 404		



QY 391 MVKDALRSFVDPVLPSPKCYLYNNHQA KCIDSFVTHCVRPFCSLIQ-----IHGHNRRARQ 446  
Db 405 -----DTFETICARSLALDVSKNHGHTNEYILQ 433  
QY 447 RDKLGHILEEFATLQDEFWTFYFNRAEKVDAAHLTMLLKQEPQ-----QHLACLGTWV 500  
Db 434 LNQMGIAGVQ-----MMQODTSRSSXNGDKSLILLNH-- 465  
QY 501 LYHNLR-----IMIQVLLSGFELELYSMHEYYIYWYLSEFLYAWLMSTLSRADGSQMAE 555  
Db 466 IYGGLEEQINWVAIRFLMLGFDLDLYSPSEYCMVYWYM--YIILWKL-----AERAR 515  
QY 556 ERIM-----EEQKQGRSSKTKKKKKVRPLSRITMSQAYQNMCMAGMFKTMVAFDMDGKV 610  
Db 516 FRVLIVVNTTEERKAKRNKEYSRDMAREDRISLWVFLKQCTCLAQGLTVMTAALRNEGMS 575  
QY 611 RKPKELDSEQVRYEHRFAPFNSVMTPPPVHYLQPKEMSDLNKYSPPPQSPELYVAASKH 670  
Db 576 LKSQGPFTENEEKFIQHFE LLQKASLPEYDAYESFSKSTSHARLD-----YLPMYEY 627  
QY 671 FQQA KMILENIP-----NPDH--EVNRILKVAKPNFVVMKLLAGGHKKESKVPPPEFDESA 723  
Db 628 FHDAQKIAKDIKVGYANDPKLAEVTGLEKVAERNIVAVNLF C-----QDRSLKVSFEFTH 683  
QY 724 HKYF 727  
Db 684 HPYF 687  
RESULT 2  
T32223  
hypothetical protein T23B12.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T32223  
R;Davidson, S.; Wohldmann, P.; Gillam, B.  
submitted to the EMBL Data Library, September 1997  
A;Description: The sequence of C. elegans cosmid T23B12.  
A;Reference number: Z21137  
A;Accession: T32223  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-814 <DAV>  
A;Cross-references: EMBL:AF022982; PIDN:AAB69936.1; GSPDB:GN00023; CESP:T23B12.4  
A;Experimental source: strain Bristol N2; clone T23B12  
C;Genetics:  
A;Gene: CESP:T23B12.4  
A;Map position: 5  
A;Introns: 24/1; 70/3; 468/2; 543/1; 616/1; 714/2  
C;Superfamily: Caenorhabditis elegans hypothetical protein T23B12.4  
Query Match 14.2%; Score 548; DB 2; Length 814;  
Best Local Similarity 24.7%; Pred. No. 6.8e-33;  
Matches 206; Conservative 131; Mismatches 302; Indels 196; Gaps 28;  
QY 18 MPEKMEKSNTNWVDITQDFEEACRELKLGELLDKLFGLPEAMSAIEMMDPKMDAGMIGN 77  
Db 1 MEGAMESPNES-EDVSKTFPKLCDGLRLGELVTTEHFRLSVDVMSAIELGEPKMDVG-VG- 57  
QY 78 QVNRKVLNFEQA IKDGTIKIKDLTLPELIGIMDTFCCLITWLEGHSLAQVFTCLYIHN 137  
Db 58 --SKHIKTLKEAISGG-LYADD--YPFQLAIMDSTLAMVVAWLEGSA LGSTVWTVNVL LSN 112  
QY 138 PDFIEDPAMKAFALGILKICDIAREKVNKA AVFEE--EDFQ-SMTYGFK----- 183  
Db 113 VTFVKHPVFPFPFASGVNLFIRNAHALINSVGNLEELPEDFNPNQLF SHQRWAPRRVVVQL 172  
QY 184 MANSVTDLRV TGM LKDVEDDMQRRVKSTRSQGBERDPEVELEHQQLAVFSRVKPTRVL 243  
Db 173 MREQVTLLGTTG-----RKWRESA FSKQA-----YDICC AVASRL EMFIML 213  
QY 244 LTVL-----IAFTKKE 254

Db 214 LEIIGLLVAPEIEDPNFDHKVHLGDFYVHEKEYEGBAENSDEKSEESTSDPTPSSEAKSE 273  
QY 255 TSAAVAEAQKLMVQAAD-----LLSAIHNSLHHGIAQN 287  
Db 274 TTEKKE-QKSNDEATDETQDNDENVGDDEDDMKQFKPNFGFASILAERLCKVTKAYA 332  
QY 288 DTTK-----GDHPIMMGFEPLVNQRLLPPTPPRYAKI IKREEMVNVFARLIDR 335  
Db 333 ETVKLGRRAPDNIDGDDYLWLGA YEPKTCIRNIPACFPRIKIVPSRQEAADWWVKCAER 392  
QY 336 IKTVCEVVNLNLHC-ILDFFCFEFSEQSPCVLSRSL LQ--TTFLVDNKKVFGTHL----- 387  
Db 393 IYHLCIVTPKTSKDLNLYLFYFARTFGQNA CVFTRSL LQICMFPVDN-----HLCGDENR 446  
QY 388 -MQDMVKDALRSFVDPVLPSPKCYLYNNHQA KCIDSFVTHCVRPFCSLIQIHGHNRRARQ 446  
Db 447 SIADAVEYSLSNCFQI LDRVSPVYKDQTAQSLYVLFNLHMSKLAITVYGSFGCNLSRQ 506  
QY 447 RDKLGHILEEFATLQDEFWTFYFNRAEK--VDAALHTMLLKQEPQRQHLACLGTWVLYHN 504  
Db 507 RDRLEMAIEDLQIHS-----YAGRLEERTDEVLLSGKWVTAKEQNYSYHSVATFVFHNL 561  
QY 505 LRIMI QYLLSGFELELYSMHEYYIYWYLSEFLYAWLMSTLSRADGSQMAEERIMEEQOK 564  
Db 562 LAIINH YFELGFRMDLYVPYEPFYIYWFIGSVQAHWMRTTLERS-----QEQLNVYQ 614  
QY 565 GRSSKKT KKK-----KV-RPLSREITMSQ-----AYQNMCMAGMFKTMVAFDMDGKV 610  
Db 615 ANPLRETKNKKLWBERCKLGEELKRRVA AHQPSVLNQIAISMISDGVVRLTVVLRKGI I 674  
QY 611 RKPKELDSEQVRYEHRFAPFNSVMTPPPVHYLQPKEMSDLNK-YSPPPPQSPELYVAASK 669  
Db 675 KMPKGGDDAEKLRFERFEPFDSLGPVVRVDRFVKSDSIDQMYEDKIET--LIDQAQK 732  
QY 670 HFQQA KMILENIPNPDHEVNR-----ILKVAKPNFVVMKLL 705  
Db 733 SFNEAREHLEKIDN-SVEQNR EMLVDAVMSFKNLLISRQLIHVAKSNIIACRVL 786  
RESULT 3  
S31288  
MAK10 protein - Yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein YEL053c  
C;Species: Saccharomyces cerevisiae  
C;Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 21-Jul-2000  
C;Accession: S31288; S30824; S50491  
R;Lee, Y.J.; Wickner, R.B.  
Genetics 132, 87-96, 1992  
A;Title: MAK10, a glucose-repressible gene necessary for replication of a dsRNA virus of  
A;Reference number: S31288; MUID:93012936; PMID:1398065  
A;Accession: S31288  
A;Molecule type: DNA  
A;Residues: 1-733 <LEE>  
A;Cross-references: EMBL:M94533; NID:g171874; PIDN:AAA34749.1; PID:g171875  
R;Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, F.  
submitted to the EMBL Data Library, February 1993  
A;Reference number: S30812  
A;Accession: S30824  
A;Molecule type: DNA  
A;Residues: 1-733 <MUL>  
A;Cross-references: GB:U18779; EMBL:L10830; NID:g603625; PID:g603626  
R;Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A;Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.  
A;Reference number: S50491  
A;Accession: S50491  
A;Molecule type: DNA  
A;Residues: 1-733 <DIE>  
A;Cross-references: EMBL:U18779; NID:g603625; PID:g603626; MIPS:YEL053c  
C;Genetics:  
A;Gene: SGD:MAK10  
A;Cross-references: SGD:S0000779; MIPS:YEL053c



A;Cross-references: GB:J05244  
R;Sahr, K.E.; Tobe, T.; Scarpa, A.; Laughinghouse, K.; Marchesi, S.L.; Agre, P.; Linnenb  
J. Clin. Invest. 84, 1243-1252, 1989  
A;Title: Sequence and exon-intron organization of the DNA encoding the alphas domain of  
A;Reference number: A45755; MUID:90009318; PMID:2794061  
A;Accession: A45755  
A;Molecule type: DNA  
A;Residues: 1-394, 'G', 396-533 <SA2>  
A;Cross-references: GB:M29983  
A;Note: the authors translated the codon GGT for residue 395 as Ala  
R;Linnenbach, A.J.; Speicher, D.W.; Marchesi, V.T.; Forget, B.G.  
Proc. Natl. Acad. Sci. U.S.A. 83, 2397-2401, 1986  
A;Title: Cloning of a portion of the chromosomal gene for human erythrocyte alpha-spectr  
A;Reference number: A23533; MUID:86205962; PMID:3458204  
A;Accession: A23533  
A;Molecule type: DNA  
A;Residues: 320-450 <LIN>  
A;Cross-references: GB:M13233; NID:g182242; PIDN:AAA53103.1; PID:g182243  
R;Curtis, P.J.; Palumbo, A.; Ming, J.; Fraser, P.; Cioe, L.; Meo, P.; Shane, S.; Rovera,  
Gene 36, 357-362, 1985  
A;Title: Sequence comparison of human and murine erythrocyte alpha-spectrin cDNA.  
A;Reference number: A91528; MUID:86083178; PMID:3000887  
A;Accession: A91528  
A;Molecule type: mRNA  
A;Residues: 1451-1687 <CUR>  
A;Cross-references: GB:M11049; NID:g338310; PIDN:AAA60569.1; PID:g553648  
R;Speicher, D.W.; Weglarz, L.; DeSilva, T.M.  
J. Biol. Chem. 267, 14775-14782, 1992  
A;Title: Properties of human red cell spectrin heterodimer (side-to-side) assembly and i  
A;Reference number: A42872; MUID:92340516; PMID:1634521  
A;Accession: A42872  
A;Molecule type: protein  
A;Residues: 7-16;46-55;680-689;1047-1056;1921-1930 <SPI>  
R;Speicher, D.W.; Davis, G.; Marchesi, V.T.  
J. Biol. Chem. 258, 14938-14947, 1983  
A;Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I domain  
A;Reference number: A92408; MUID:84087888; PMID:6654896  
A;Accession: A02965  
A;Molecule type: protein  
A;Residues: 7-601 <SPC>  
R;Speicher, D.W.; Davis, G.; Yurchenco, P.D.; Marchesi, V.T.  
J. Biol. Chem. 258, 14931-14937, 1983  
A;Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain and  
A;Reference number: S13138; MUID:84087887; PMID:6654895  
A;Accession: S13138  
A;Molecule type: protein  
A;Residues: 7-92, 'X', 94-96, 110-151;317-342;345-366, 'Z', 368-370, 'X', 372-373;387-434;452-4  
R;Lusitani, D.M.; Qtaishat, N.; LaBrake, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fung,  
J. Biol. Chem. 269, 25955-25958, 1994  
A;Title: The first human alpha-spectrin structural domain begins with serine.  
A;Reference number: A38928; MUID:95014412; PMID:7929303  
A;Contents: annotation; determination of structural domain  
C;Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal superstr  
s of approximately 106 residues each.  
C;Genetics:  
A;Gene: GDB:SPTA1  
A;Cross-references: GDB:119601; OMIM:182860  
A;Map position: 1q21-1q21  
C;Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/  
C;Keywords: actin binding; cytoskeleton; duplication; EF hand; erythrocyte; heterodimer;  
F;52-156/Domain: spectrin/dystrophin repeat homology <SP1>  
F;157-262/Domain: spectrin/dystrophin repeat homology <SP2>  
F;263-368/Domain: spectrin/dystrophin repeat homology <SP3>  
F;369-474/Domain: spectrin/dystrophin repeat homology <SP4>  
F;475-580/Domain: spectrin/dystrophin repeat homology <SP5>  
F;581-685/Domain: spectrin/dystrophin repeat homology <SP6>  
F;686-791/Domain: spectrin/dystrophin repeat homology <SP7>  
F;792-897/Domain: spectrin/dystrophin repeat homology <SP8>  
F;898-983/Domain: spectrin/dystrophin repeat homology #status atypical <SP9>  
F;984-1031/Domain: SH3 homology <SH3>  
F;1081-1181/Domain: spectrin/dystrophin repeat homology <SP10>  
F;1182-1287/Domain: spectrin/dystrophin repeat homology <SP11>  
F;1288-1393/Domain: spectrin/dystrophin repeat homology <SP12>

F;1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>  
F;1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>  
F;1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>  
F;1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>  
F;1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>  
F;1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>  
F;2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>  
F;2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>  
F;2270-2302/Domain: calmodulin repeat homology <BF1>  
F;2313-2345/Domain: calmodulin repeat homology <BF2>  
  
Query Match 3.5%; Score 134; DB 1; Length 2429;  
Best Local Similarity 19.4%; Pred. No. 0.37;  
Matches 175; Conservative 142; Mismatches 307; Indels 280; Gaps 48;  
  
QY 13 GWELSMPEKMEKSNWN---VDITQD---FEEACRELKLGELLHDKLFGLFEMSAIEM 65  
Db 1572 GNEEAMKEQLEQLKEHWDHLLERTNDKGKLNESRQQRNTSIRDFFFWLSEATLLAM 1631  
  
QY 66 MDPKMDAGMIGN-----QVNRKVLNFEQAIDGTIKIKDLTLPELIGIMDTCFCLITW 119  
Db 1632 KDQARDLASAGNLKXHQLLEREMLAREDAKDLNLTAE DL----- 1672  
  
QY 120 LEGHSLAQTVFTCLYIHNPDFI---EDPAMKAFALGILKICDIAREKVKA-AVF----- 170  
Db 1673 -----LSSGTF-----NVDQIVKKDNVNRKF-LNVQELAAAHHEKLEAYALFQFFQD 1720  
  
QY 171 --EEEDF-----QSMTYGFKMANSVTDL-----RVTGMLKDVED-----DMQRRVK 209  
Db 1721 LDDEESWIEEKLIRVSSQDYG-RDLQGVQNLKXHKRLEGELVAHEPAIQNVLDMAEKLK 1779  
  
QY 210 STRSRQGEERDPEVEL-----EHQQCLAVFSRVKTRVLLTV---LIAFTKKETSAVA 259  
Db 1780 D-KAAVGQE---EIQLRLAQFVEHWEKLEKAKARGLKLEESLEYLQFMQNAEEEEAWIN 1835  
  
QY 260 EAQKLMVQ--AADLLSAI-----HNSLH-----HGIAQNDTTKGDPHPIMMGFEPLVNQ 306  
Db 1836 EKNALAVRGDCGDTLAATQSLLMKHEALENDFAVHETRVQNVCAQG-----EDILNK 1887  
  
QY 307 RLLPPTFPYAKIIKREEMVNYFARLIDRIKTVCEVVNLTNLHCILDF-FCFSEQSPCV 365  
Db 1888 VLHEES-----QNKIEISSKIEALNEKTPSLAKAIAAWKQLQLEDDYAFQEFNWKADV 1939  
  
QY 366 LSRSLLOTTFVLVDNKKVFGTHL-----MQDMVKDALRSVD---PPVLSPKCVLYN- 413  
Db 1940 EAWIADKETSLKTGN--GADLGDFLTLLAKQDTLDASLOSFOQERLPEITDLKDKLISA 1997  
  
QY 414 NHQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQ---DEFTTFYF- 469  
Db 1998 QHNQSKAIEERYAALLKRWEQLLEASAVHRQK-----LLEKQLPQKAEDLFVEFAHK 2050  
  
QY 470 -----NRAEKVDAAL-----HTMLLKQ--EPQRQH-----LAC 495  
Db 2051 ASALNNWCEKMEENLSEPVHVCVSLNEIRQLQKDHEDFLASLARAQADFKLLELDQOIK 2110  
  
QY 496 LG-----TWVLYHNLRIMIQYL-----LSGFEL-ELYSMEHYVYIY 530  
Db 2111 LGVPSSPYTWLTVEVLERTWKHLSDIIEEREQELQKBEARQVKNFEMCQEFQNASTFLQ 2170  
  
QY 531 WYLSEFLYAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREIT----- 585  
Db 2171 WILETRAYF-----LDGSLKETGTLESQLEANK---RKQKEIQAMKRQLTKIVDL 2218  
  
QY 586 -----MSQAYQ---NMCAGMFKTMVAFDMDGKVRKPKFELDSE 620  
Db 2219 GDNLEDALILDIKYSTIGLAQQWDQLYQLGLRMQHNLEQQIQAKDIKGVSEETLKEFST- 2277  
  
QY 621 QVRYEHRFAPFNSVMTPPPVHYLQFKE-MSDLNKYSPPPQSPEL----- 663  
Db 2278 --IYKH-----FDENLTGLRTH-KEFRSCLRGLNYLPMVEEDEHEPKFEKFLDAVDPRK 2330  
  
QY 664 -YVAASKHFQQAAMI-----LENIPNDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEF 719  
Db 664 -YVAASKHFQQAAMI-----LENIPNDHEVNRILKVAKNFVVMKLLAGGHKESKVPPEF 719



Db 2331 GYVSLDY--TAFLLDKESENIKSSDEIENAFQALAEGKSYTK-----EDMKQALTPEQV 2384

QY 720 DFSA 723

Db 2385 SFCA 2388

RESULT 6

S20106

hypothetical protein 2 - slime mold (Dictyostelium discoideum) retrotransposon DRE

C;Species: Dictyostelium discoideum

C;Date: 19-Mar-1997 #sequence revision 25-Apr-1997 #text\_change 01-Dec-2000

C;Accession: S20106; S30209; S30211; S30207; S25646

R;Marschalek, R.; Hofmann, J.; Schumann, G.; Goesseringer, R.; Dinger mann, T.

Mol. Cell. Biol. 12, 229-239, 1992

A;Title: Structure of DRE, a retrotransposable element which integrates with position sp

A;Reference number: S20105; MUID:92107173; PMID:1309589

A;Accession: S20106

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1152 <MAR>

A;Cross-references: EMBL:X57034

R;Marschalek, R.; Hofmann, J.; Schumann, G.; Dinger mann, T.

Nucleic Acids Res. 20, 6247-6252, 1992

A;Title: Two distinct subforms of the retrotransposable DRE element in NC4 strains of Di

A;Reference number: S30206; MUID:93117097; PMID:1335570

A;Accession: S30209

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-87,'RNR','91','S' <MA2>

A;Cross-references: EMBL:X69170; NID:g7259; PIDN:CAA48919.1; PID:g7261

A;Experimental source: retrotransposon DRE, subtype b; strain AX2

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992

A;Accession: S30211

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-87,'RNR','91','S' <MAW>

A;Cross-references: EMBL:X69171; NID:g7262; PIDN:CAA48921.1; PID:g7263

A;Experimental source: retrotransposon DRE, subunit b; strain AX2

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992

A;Accession: S30207

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-87,'XQ' <MA3>

A;Cross-references: EMBL:X68231

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992

C;Genetics:

A;Mobile element: retroposon DRE

Query Match 3.4%; Score 132.5; DB 2; Length 1152;

Best Local Similarity 17.6%; Pred. No. 0.17;

Matches 130; Conservative 120; Mismatches 267; Indels 223; Gaps 37;

QY 18 MPEKMEKSNTNWVDITQDFEEACR-----ELKIGELLHDKLFGLFEMSALIEMDPK 69

Db 370 IPQILDPSNNCLVTKHEDILEVARRYENLYQKRECNEETHHELKTFNKRIEQKILD-- 427

QY 70 MDAGMIGNQVNRKVLNFEQAIKDGTIKIDLTLPPELIGIMDTFCCLITWLEGH-SLAQT 128

Db 428 -----EINQPIEGYE--IRLGIKIQEGKAPGKDGLLPTFY-----KNHINEILP 470

QY 129 VFTCLYIHNPFIEDPAMKAFALGILKICDIAREKVNKAAVFEEEFQSMTYGFKVANSV 188

Db 471 IISKLYNH---FWNTTIPKDFKQGIL--ITLYKNKGDPNNLDNYRPIITLLNVVDYKIYSKI 525

QY 189 TDLRVTGMLKDVEDDMQ-----RRVKSTRSRQGEERDPEV-- 223

Db 526 INNRILKLLNKIISPFQTGFVPRRLLDHNIITLNSTIEIKREINTK----EDMEPIITF 581

QY 224 -----ELEHQOCLAVFSRVKF--TRVLLTVLIAFTKKETSAAVEAQKLMVQAADLLSA 274

Db 582 YDFEKAFDISISHNAILRTLHLKPLPKMVLTTMNLNSETSVY----- 625

QY 275 IHNSLHHGIAQNDDTTKGD--HPIMMGF-----EPLVNQRLLPPTFPFYAKIIKREE-- 324

Db 626 INNSLSKSFTSKRGTKQGDPISPITFALVVECMATTIINDRCINGVTKETIKILOFADDT 685

QY 325 -----MVNYFEARLIDRIKTVCEV-----VNLTNLHCILDFCFEFSEQSPCVLSRSLQTT 374

Db 686 ATIAYNFMDHFL-MNEWIKKFCQATSAKINQTKSCI-----TFKWN-----TRTLY--T 732

QY 375 FLVDNKKVFGTHLMQDMVKDALRSFVDPFVLSPKCYLYNNHQAQDCIDSFVTHCVRPFC- 433

Db 733 VIKSNERYLGFDFNNKGKISKINTISDN--IRAKLVTWNSTSS-----TYMGRLLIMA 782

QY 434 ---SLIQIHGH---NRARQDKLGHILEEFATLQDEFMTFFNRAEKVDAALHTMLLKQE 487

Db 783 KTYALSQLTFTHTYINTTPQHN-----SIENNIVKFVFNTKSKNSLSLQ----- 825

QY 488 PQQHACLGTWVLYHNLRIMIQLYLLSGFELELYSMHEYIYYIYVYLSEFLYAWLMSTLSR 547

Db 826 -RRQN-----NYINGGLNLWNLKTRELAQKAWLFEYLH-----QR 860

QY 548 ADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNWCAGMFKMTMVAFDMD 607

Db 861 VSNTPSSYIKLWEEELKNNNNKTTTKQ--NQLQHLWQCKQAWT----- 902

QY 608 GKVRKPKFELDSEQVRYEHRFAP-----FNSVM-TPPPVHYLQFKEMSDLNKYSPPQSP 661

Db 903 -QLKTPQ----NKQTHYEH--LPKLKKIYEDMMTTQSPEH-----NKFIPTPGQK 945

QY 662 ELYVAA-SKH--FQQAQMIL 678

Db 946 EIMTKINSKHLPPFKEIKKII 965

RESULT 7

E71810

type III restriction enzyme - Helicobacter pylori (strain J99)

C;Species: Helicobacter pylori

A;Variety: strain J99

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999

C;Accession: E71810

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; J

Nature 397, 176-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: E71810

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-969 <ARN>

A;Cross-references: GB:AE001563; GB:AE001439; NID:g4156027; PIDN:AAD06988.1; PID:g4156025

A;Experimental source: strain J99

C;Genetics:

A;Gene: res\_2

Query Match 3.3%; Score 127; DB 2; Length 969;

Best Local Similarity 17.3%; Pred. No. 0.34;

Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;

QY 35 DFEEACRELKIGELLHDKLFGLFEMSALIEMDPKMDAGMIGNQV-----NRK 82

Db 9 DYQEQCR-----DQILGVFK---GIDLREPENDIQRIANPVFEIGAICKDLLENIE 56

QY 83 VLNFEQAIKDGTIKIDLTLPPELIGIMDT-----CFCCCLITWLEGSLAQTFTCLY---- 134

Db 57 NLRSKQKITQGSVGINQSLNCDI--LNETGTGKTFCFL-----ECVYALHK 100

QY 135 -IHNPDFIEDPAMKAFALGILKICDIAREKVNKAAVFEEEFQSMTYGFKVANSVTDLRV 193

Db 101 NYHLSKFIVLAPSNAIKGLVLSKSIETRE-----FFKSEYSNTHL-- 140

QY 194 TGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFTREVLLTV----- 246

Db 141 -----ESYEDVERFILAS-----NHKCCVLVMTFSAPNKEKNTINQSCLEN 181

QY 247 --LIAFTTKETSAAVAEQKLMV-----QAADLLSAIHNSLHHGIAQNDTTK 291  
Db 182 TNLFNKAKSYMQUALASMRPIVIMDEPHRFLGDKTKKYLEQLNAL-ITLRFGATFKDDY-- 238  
QY 292 GDHPIMMGFEPLVNQRLPPTFPYAKIIKREEMVNYFARLID----- 334  
Db 239 --HNLIIYALD---SKKAFDCALVKSISVASVGESDEYFELKEANKKQNEATINYTTLEN 293  
QY 335 -----RIKTVCEVVNLTNLHCILDFFCFESEQSPCVLSRSLTQTTLVDNKKVFGTHLMQ 389  
Db 294 KIQSVKVKTHDNLGALTHISALEDIYVEKITKTEI---RFLNGFNLLLDQKEPF-SHLL 349  
QY 390 D-----MVKDALSRLFV-----PPVLSPKCYLYNNHQKDCIDSFVTHCVRP-----FC 433  
Db 350 GEQEVMLKEAIKSHFEREEGLFKKGIKALCMVFING-----VNSYLSENEKPAKLALIFE 404  
QY 434 SLIQIHGHNRRARQDKLGHILLEEAFATLQDEFMTFFYNRAE---KVDAALHTMLLKQEPQ 489  
Db 405 KLYQ-----QKLEEVLLK---PLDENYRAYLERTKDAILKVHGGYFAKSKKESDE 451  
QY 490 RQHLACLGTWVLYHNLRIMIQVLLSGFELELYSMHEYYIYWYLSEFLYAW---LMSTL 545  
Db 452 VQVIA-----LILKEKELLSFSDSL---RFIFSQWALQE---GWDNPNVMTIC 494  
QY 546 SRADGSQMAEERIMEEQGRSSKTKKKKKVR----- 578  
Db 495 Klapss-----SNITKLOQIGRGLRLAVNDKGERITKEHADDFVNLVVIVPQEGDFVG 550  
QY 579 PLSREITMSQAYQNMCMAG-----MFKTMVAFDMDGKVRKPKPFELDSEQV 622  
Db 551 AIQOEISEHSLIKQVFSAEELKSGMVKKGYGVLFETLEGLGFGKTDDENFKLTLNQ- 609  
QY 623 RYEHFAPFNSVMTPPP-----VHYLQFKEMSDLNKYSPPPQSPELYVAASKHPQQA 675  
Db 610 -----NEFLKKEPELEKLDKDYLDFEKLKDFLK-----DRLIG---HPR--- 646  
QY 676 MILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKKESKVPPEFD 720  
Db 647 -----VRNKNERKTEKINKENFKKFETLWAGLNHQARIAYAID 686  
RESULT 8  
S48244  
NMD2 protein - Yeast (Saccharomyces cerevisiae)  
N;Alternate names: IFS1 protein; protein YHR077c  
C;Species: Saccharomyces cerevisiae  
C;Date: 28-Oct-1994 #sequence revision 10-Feb-1995 #text\_change 23-Mar-2001  
C;Accession: S48244; S46815; S64648; S64738  
R;He, F.; Jacobson, A.  
submitted to the EMBL Data Library, September 1994  
A;Description: Identification of a novel component of the nonsense-mediated mRNA decay p  
A;Reference number: S48244  
A;Accession: S48244  
A;Molecule type: DNA  
A;Residues: 1-1089 <HEF>  
A;Cross-references: EMBL:U14974; NID:G5555938; PID:G5555939  
R;Favell, T.  
submitted to the EMBL Data Library, June 1994  
A;Description: The sequence of S. cerevisiae cosmid 9205.  
A;Reference number: S46795  
A;Accession: S46815  
A;Molecule type: DNA  
A;Residues: 'MYQQ', 3-1089 <FAV>  
A;Cross-references: EMBL:U10556; NID:G500825; PID:G500836; MIPS:YHR077c  
R;Lee, S.I.; Umen, J.G.; Varmus, H.E.  
Proc. Natl. Acad. Sci. U.S.A. 92, 6587-6591, 1995  
A;Title: A genetic screen identifies cellular factors involved in retroviral -1 frameshi  
A;Reference number: S64648; MUID:95327692; PMID:7604038  
A;Accession: S64648  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1089 <LEE>

A;Cross-references: EMBL:U28158  
R;Varmus, H.E.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S64738  
A;Accession: S64738  
A;Molecule type: DNA  
A;Residues: 'MYQQ', 3-1089 <VAR>  
A;Cross-references: EMBL:U28158; NID:G967212; PID:G967213  
C;Genetics:  
A;Gene: SGD:NMD2; IFS1  
A;Cross-references: SGD:S0001119; MIPS:YHR077c  
A;Map position: 8R  
A;Introns: 2/3

Query Match 3.3%; Score 127; DB 2; Length 1089;  
Best Local Similarity 17.6%; Pred. No. 0.4;  
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;

QY 24 KSNTNWV-----DITQDFEEACRELKLGELL-----HDKLFGLF 57  
Db 35 KRNTGFIKKLKKGFVKGSSESLKDLSEASLEKYLSEIIVTVTECLLVNKNDDVIAAV 94  
QY 58 EAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIGDTIKIKDITLPELIGIMDTFCCL 116  
Db 95 EIISGLHQRFNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR 136  
QY 117 ITWLEGHSLAQVFTCLYIHN-----PDFIE-----DPAMKAFALGILK- 155  
Db 137 ITRVKN- --LRVFTELYLVGVFTLDDIESKDAIFNLOKTKGRKDPPLFSILREILNY 193  
QY 156 -----ICDIAREKVNK-AAVFEED--FQSMTYGFKMANSVTDL-----RVTGM 196  
Db 194 KFKLGFTTTIATAFIKKFAPLFRDDDNWDLLIYDSKLKALQSLFKNFIDATFARATEL 253  
QY 197 LKDVEDDMQRRVKSTRSQGEERDPEVELEHQCLAVFSRVKTRVLLTLVLIATFKKETS 256  
Db 254 HKKV-NKLQREHQKQIRTKGLRDEYVE-EYDKLLPIRKF-----TS 295  
QY 257 AVAEAQKLMVQAADLLSAIHNSLHHGIAQNDDTTKGDPHIMMGFEPLVNQRLPPTFPY 316  
Db 296 AITLGEFFKLEIPELEGASNDL-----KETA-----SPMITNQILPPN--- 334  
QY 317 AKIIKREEMVNYFARLIDRIKTVCEV-----VNLTNLHCILDFFCFESEQSPCVL 366  
Db 335 QRLWENEDTRKFYEILPDISKTVESQSSKTEKDSNVNSKNINL---FFTDL-EMADC-- 388  
QY 367 SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA 417  
Db 389 -----KDIIDDLNRYWSSYLDNKATRNRILKFFMETQDWSKLPVYS-RFIATNSKYM 440  
QY 418 KDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFMTFFYNRAEKVDA 477  
Db 441 PEIVSEFINYLDNGFRS--QLHS-----NKNVKN 469  
QY 478 ALHTMLLKQEPQROHLACLGTWVLYHNLRIMIQVLLSGFELELYSMHEYYIYWYLSEFL 537  
Db 470 IFFSEMIK-----FQLIPSFMIHFKIRTLIMYQVNNVEILTV-----LLEHSGKFL 517  
QY 538 YAWLMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKVRPLSREITMSQAYQNMCMAGM 597  
Db 518 -----LNKPEYKELMEKMW-----QLIKDKKNDRQLN--MNMKSALENIITLL 558  
QY 598 FKTMV 602  
Db 559 YPPSV 563

RESULT 9  
T16111  
hypothetical protein F20D12.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
C;Accession: T16111







Db 1082 LEAPQMYWNTLRTVCDKWIRMKVEMDARRSPMEHE-----NKEWTLRTELLAARPQMP 1137

QY 638 PPVHYLQFKEMSDLN-----KYSPPQSPELYVAASKHFQQAkmilenIPN-----PDHE 687

Db 1138 SSLRQLBEQPHREVQLAFNAKWSGSKSEITVNAQLEQSTEQKXFIRNIERYKGIPEYE 1197

QY 688 VNRILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSAHKYFPVVK 731

Db 1198 L--LIKAARLNQVNV-----VSEYKLTPOSEYTFSRIFDLIK 1232

RESULT 13

T13057

KIAA0729 protein - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C;Accession: T13057; T08741

R;Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Miyajima, N.; Tanaka, A.; Kotani, H. DNA Res. 5, 277-286, 1998

A;Title: Prediction of the coding sequences of unidentified human genes. XI. The completed sequence of the cDNA for human proteinase 3.

A;Reference number: Z17595; MUID:99087487; PMID:9872452

A;Accession: T13057

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1196 <NAG>

A;Cross-references: EMBL:AB018272; NID:g3882178; PIDN:BAA34449.1; PID:g3882179

A;Experimental source: brain

R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, March 1999

A;Reference number: Z16471

A;Accession: T08741

A;Molecule type: mRNA

A;Residues: 279-1196 <WAM>

A;Cross-references: EMBL:AL050092

A;Experimental source: adult uterus; clone DKFZp586G0518

C;Genetics:

A;Note: KIAA0729; DKFZp586G0518.1

Query Match 3.1%; Score 119; DB 2; Length 1196;

Best Local Similarity 19.3%; Pred. No. 1.8;

Matches 124; Conservative 99; Mismatches 235; Indels 186; Gaps 31;

QY 97 IKDLTLPeligIMDTCFCCLITWLEGHSLAQTVFTCLYIHNPFDIEDPAMKAFALGILKI 156

Db 563 MREELEDIKQFKKTTISCYLRCLDGRSCWTLIS-----AFRILLJE 604

QY 157 CDIAREKV--NKAAVFEEEDFQSMTYGFKVANSVTDLRVTGMLKDVEDDMQRRVKSTRSR 214

Db 605 SDEDRLLVFNRLGILIMTESFNTL---HMMYHEATACHVTGDLVELLSIFLSVLKSTRPY 661

QY 215 QGEERDPEVELEHQCLAVFSRVKFTRVLLTVLIAFTKKE-TSAVAEAKQ--LMVQAADL 271

Db 662 LQRKQVKQALIQWE-----RIEFAHKLLTLNLSYSPPELRNACIDVLKELVLLSPHDF 715

QY 272 LSAI-----HNSLHHGIGIAQNDDTTKGDPIMMG--FEPLVNQRL-----PP----- 311

Db 716 LHTLVPEFLQHNHCTYHH-----SNIPMSLGPYFPCCRENIKLIGGKSNIRPPRPPEL 765

QY 312 ---TFPRYAKIKREEMVNYFARLIDRIKTVCEVVNLTNLHCILDFCE-FSEQSPCVLS 367

Db 766 NMCLLPMTWVETSGKDDV-YDRMLLDYFFSHQFIHLL---CRVAINCEKFT-----814

QY 368 RSLLOTTFLVDNKKV-----FGTHLMQD---MVKDALRSFVDPVLSP--KC-----409

Db 815 -TLVKLSVLVAVEGLPLHLALFPKLWTELCQTQSAMSKNCIKLLCEDPVFAEYIKCILMD 873

QY 410 ---YLYNNHQAQDCIDSFVTHCVRPFCSLIQIGHNRRARQDKLG----HILEEFATLQD 462

Db 874 ERTFLNNN-----IVYTFMTH-----FLLKVQSQVFSEANCANLISTLITNLISQYQNLS 924

QY 463 EFMTFYFNRAE--KVDAALH-----TMLLKQBPQRQHLACLGTWLVHNLNIRIMIQYLLS 514

Db 925 DFS----NRVEISKASASINGDLRALALLSVHTPKQ-----957

QY 515 GFEELEYSMHEYYIYWYLSEELYAWLMTLSRADGSQMAEERIMEEQKGRSSKTKKK 574

Db 958 -----LNPALIPTLQELLSKRCCTLQORNSLQEQEAKER---KTXDD 996

QY 575 KKVRLSREITMSQAYQNM--CAGMFKTMVAFMDGMGKVVRKPKFELDSEQVRYEHRFAPEN 632

Db 997 EGATPIKRRRVSSDEEHTVDSICISDMKTETR-----EVLTPSTSDNE-----TRDS 1043

QY 633 SVMTPPPVHYLQFKEMSDLNKYS---PPPQSPELYVAASKHFQO 673

Db 1044 SIIDPCTEQDLFSPENS SVKEYRMEVPSSFSEDSMNIRSQAHEE 1087

RESULT 14

T02576

hypothetical protein At2g39260 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T16B24.10

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001

C;Accession: T02576; B84815

R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S. submitted to the EMBL Data Library, August 1998

A;Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.

A;Reference number: Z14679

A;Accession: T02576

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1029 <ROU>

A;Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402679

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: B84815

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1029 <STO>

A;Cross-references: GB:AE002093; NID:g3402679; PIDN:AAC28982.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g39260; T16B24.10

A;Map position: 2

A;Introns: 72/3; 104/3; 147/2; 205/2; 232/3; 330/3; 396/3; 430/3; 514/3; 587/3; 644/3; 70

Query Match 3.1%; Score 118.5; DB 2; Length 1029;

Best Local Similarity 18.8%; Pred. No. 1.6;

Matches 87; Conservative 82; Mismatches 132; Indels 163; Gaps 23;

QY 21 KMEKSNTNWV---DITQDFEEACRELKLGELLHDKLFGLFEMS-AIEMMDPKMDAGMI 75

Db 108 QMEKENAKLVNAKGELSEDSASSYEKLRS---YDHLRYNISSLAHALDMQPPVMP----160

QY 76 GNQVNRKVLNFEQAIKDGTIKIKDLTLPeligIMDTCFCCLITWLEGHSLAQTVFTCLYI 135

Db 161 -EDGTTRLTAGDEASPGT--VKDTSVPEPI-----WDEDT--KTFYECL--201

QY 136 HNPFDIEDPAMKAFALGIL-----KICDIAREKVNKAA---VFEEEDFQSMTYGFK 183

Db 202 -----PDLRAFVPAVLLGEAPKSNQSAKAKEKLSSESSEVVENQQTTEVSA 253

QY 184 MANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQ-----QCLAVFSRVKF 239

Db 254 DSASMDDRSNAEQPKKEEVEKEKADTKKEKGKEKDSKKMEHEKEKGSLDV---ANF 310

QY 240 TRVL-----LTVLIAFTKKETSAVAEAKLMVQAADLLSAIHNSLHHGIGA 285

Db 311 ERLQLRPGCVSRDLIDQLTVEYCYLNSKTN-----RKKL VKA--L FNVVPTSL-----357

QY 286 QNDTTKGDPIMMGFEPLVNQRLPPPTFPYAKIKREEMVNYFARLIDRIKTVCEVVNL 345

Db 358 -----ELLAYSRMV-----A 368

QY 346 TNLHCILDFFCFSEQSPCVLSRLQTTF--LVDNK-----KVFGTHLMQDMVKDAL 396

Db 369 TLASCMKDI-----PSMLVQ-MLEDEFNSLVHKKDQMNIRFIGELCKFEK- 418

QY 397 RSFVDPVPLSPKCYLYNNHQAKDCIDSFVTHCVRPFCSLIQIHG 440

Db 419 ---IVPAGLVFSC-----LKACLDFTFHHNIDVACNLLLETCG 452

RESULT 15

S02041

dystrophin, muscle - chicken

N;Alternate names: duchenne muscular dystrophy protein

C;Species: Gallus gallus (chicken)

C;Date: 07-Sep-1990 #sequence\_revision 27-Jun-1994 #text\_change 16-Jul-1999

C;Accession: S02041; S02013; S71487

R;Lemaire, C.; Heilig, R.; Mandel, J.L.

Nucleic Acids Res. 16, 11815-11816, 1988

A;Title: Nucleotide sequence of chicken dystrophin cDNA.

A;Reference number: S02041; MUID:89098331; PMID:3062582

A;Accession: S02041

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-3660 <LEM>

A;Cross-references: EMBL:X13369; NID:g63369; PIDN:CAA31746.1; PID:g63370

A;Note: 1869-His, 1885-Arg, and sequences lacking 1171-Met were also found

R;Lemaire, C.; Heilig, R.; Mandel, J.L.

EMBO J. 7, 4157-4162, 1988

A;Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding and

A;Reference number: S02013; MUID:89210800; PMID:3072195

A;Accession: S02013

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-3573,'HA', 3576-3660 <LEM2>

R;Heilig, R.; Lemaire, C.; Mandel, J.L.

Nucleic Acids Res. 15, 9129-9142, 1987

A;Title: A 230kb cosmid walk in the Duchenne muscular dystrophy gene: detection of a con

A;Reference number: S09071; MUID:88067745; PMID:2825128

A;Accession: S71487

A;Molecule type: DNA

A;Residues: 222-281 <HEI>

C;Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl

C;Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro

C;Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystro

C;Keywords: actin binding; calmodulin binding; cytoskeleton; leucine zipper; membrane-as

F;18-237/Domain: alpha-actinin actin-binding domain homology <ACT>

F;253-327/Region: hinge

F;340-449/Domain: spectrin/dystrophin repeat homology <SP01>

F;450-558/Domain: spectrin/dystrophin repeat homology <SP02>

F;560-669/Domain: spectrin/dystrophin repeat homology <SP03>

F;670-719/Region: hinge

F;720-830/Domain: spectrin/dystrophin repeat homology <SP04>

F;838-936/Domain: spectrin/dystrophin repeat homology <SP05>

F;940-1047/Domain: spectrin/dystrophin repeat homology <SP06>

F;1049-1156/Domain: spectrin/dystrophin repeat homology <SP07>

F;1158-1265/Domain: spectrin/dystrophin repeat homology <SP08>

F;1267-1369/Domain: spectrin/dystrophin repeat homology <SP09>

F;1374-1479/Domain: spectrin/dystrophin repeat homology <SP10>

F;1480-1570/Domain: spectrin/dystrophin repeat homology #status atypical <SP11>

F;1572-1678/Domain: spectrin/dystrophin repeat homology <SP12>

F;1680-1784/Domain: spectrin/dystrophin repeat homology <SP13>

F;1787-1877/Domain: spectrin/dystrophin repeat homology #status atypical <SP14>

F;1878-1984/Domain: spectrin/dystrophin repeat homology <SP15>

F;1986-2103/Domain: spectrin/dystrophin repeat homology <SP16>

F;2105-2211/Domain: spectrin/dystrophin repeat homology <SP17>

F;2213-2319/Domain: spectrin/dystrophin repeat homology <SP18>

F;2323-2419/Domain: spectrin/dystrophin repeat homology <SP19>

F;2420-2467/Region: hinge

F;2468-2574/Domain: spectrin/dystrophin repeat homology <SP20>

F;2576-2683/Domain: spectrin/dystrophin repeat homology <SP21>

F;2685-2799/Domain: spectrin/dystrophin repeat homology <SP22>

F;2801-2928/Domain: spectrin/dystrophin repeat homology <SP23>

F;2930-3037/Domain: spectrin/dystrophin repeat homology <SP24>

F;3038-3075/Region: hinge

F;3052-3089/Domain: WW repeat homology <WW1>

F;3079-3357/Region: cysteine-rich

F;3481-3502/Region: leucine zipper motif

F;3547-3568/Region: leucine zipper motif

Query Match 3.1%; Score 117.5; DB 1; Length 3660;

Best Local Similarity 17.7%; Pred. No. 11;

Matches 132; Conservative 113; Mismatches 229; Indels 273; Gaps 35;

QY 13 GWELSMPEKMEKSNNTNWVDIT----- 33

Db 1646 GKESLVEDKLSLLNSNWIAVTSRAEEWLNLMEYQKMEAFDQKVANVTTWIYRAEILLD 1705

QY 34 -----QDFEEACRELKLGEL--LHDKLFGLFEAMSAIEMM-----DPKMDAG 73

Db 1706 ESDKQKPQKEETLKRK-AELNDMHPKVDVSRD--QAVDLMTNRGDHCRCVIEPKL--- 1759

QY 74 MIGNQVNRKVLNFEQAIKDGTIKIKDLTLPGLIGIMDTFCCLITWLEGHSLAQTVFTCL 133

Db 1760 ---SELNHRFAAISQRIKSG----- 1776

QY 134 YIHNPDFIEDPAMKAFALGILKICDIAREKVNKAAVFEEDF-----QSMTY 180

Db 1777 ----KPFIPLKELEQDFDIQKLLPLEVEIQGVNLMKEEDFNKDMSEDDSTVVKELLQR 1832

QY 181 GFKMANSVTDLRVTCMLKDVEDDDMQRRVKSTRSQGEERDPEVELEHQOCLAVFSRVKFT 240

Db 1833 GDTLQKRITDERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEISHOW----YQYKQA 1888

QY 241 RVLLTVLIAFTKKETSAV--AEAQKLMVQAADL-----LSAIHNSLHHGICQANDTTKG 292

Db 1889 DDLMNTWLDIEKKLASLPDHKDEQKLKEIGGELEKKEDLNANVR-----QAERLSKDG 1942

QY 293 DHPIMMGPEPLVNRLLPPTFPYAKIIRREEMVNVPFARLIDRIKTVCEVNVNLNLHCIL 352

Db 1943 -----AAKAVEPTLVQLSK--RWRDFESKPAQF-----RRLNYAQIQTVL 1980

QY 353 DFFCEFSQSPCVLSRL-LQTTFLVDNKKVFGTHLMQDM--VKDALRSFVDPVPLSPKC 409

Db 1981 -----EDTTFVMTESMTVETTVPSTYLAEILQLLQALSEVEERLNS---PVL--- 2025

QY 410 YLYNNHQAKDCIDSF-VTHCVRPF--CSLIQIHGHNARQRDKLGHIL---EEFATLQD 462

Db 2026 -----QAKDCEDLLQBECLNKKDC-LGRLOGHIDIHSKKTPALQSATPRETANIQD 2078

QY 463 EFMTFYFNRAEKVDAALHTWLLKQEPQROHLACLGTFWVLYHNLIRIMIQYLLSGFELELYS 522

Db 2079 K-LTQLNSQWEKVNKMVYDRDQARFDKSKE-----KWRLFH-----CEMKS 2117

QY 523 MHEYIIYIYWLSEFLYAWLMST---LSRA-----DGSQMAEERIMEEQKGRSSKKTKKK 574

Db 2118 FNE-----WLTETEKLKLSRAQIEAGDVGHVKTQKQFLQELQDGIGRQQTIVVK 2163

QY 575 KKVRLPSREITMSQAYQNMCAGMFKTMVAFMDGKVRKPKF-----ELDSEQVRYEH 626

Db 2164 -----TLNVTGEETIEQSSAA-----DANVLKEQLGNLNRWQEICRQLVEKRRK 2207

QY 627 RFAPFNSVMTPPPVHYLQFKEMSDLNK 653

Db 2208 RIEEEKNILS-----EFQE--DLNK 2225

Search completed: August 10, 2004, 21:52:01

Job time : 25 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OMprotein - protein search, using sw model

Run on: August 10, 2004, 21:47:26 ; Search time 13 seconds  
(without alignments)  
2935.956 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MVMKASVDDDDSGWELSNPE.....KVPPEFDPSAHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	270.5	7.0	733	1 MK10_YEAST	Q02197 saccharomyc
2	134	3.5	2418	1 SPCA_HUMAN	P02549 homo sapien
3	127	3.3	733	1 SAB2_HUMAN	Q9upw6 homo sapien
4	127	3.3	733	1 SAB2_MOUSE	Q8vi24 mus musculu
5	127	3.3	1089	1 NMD2_YEAST	P38798 saccharomyc
6	125.5	3.3	1603	1 VIT3_CAEEL	Q9n4j2 caenorhabdi
7	125.5	3.3	1603	1 VIT4_CAEEL	P18947 caenorhabdi
8	121.5	3.2	962	1 IF3A_MAIZE	Q9xhr2 zea mays (m
9	121.5	3.2	1603	1 VIT5_CAEEL	P06125 caenorhabdi
10	117.5	3.1	3660	1 DMD_CHICK	P11533 gallus gall
11	116	3.0	736	1 BAC1_HUMAN	O14867 homo sapien
12	115	3.0	1187	1 A11A_MOUSE	P98197 mus musculu
13	114	3.0	1679	1 YIO9_YEAST	P40457 saccharomyc
14	113.5	2.9	1957	1 SPOF_SCHPO	Q10411 schizosacch
15	113	2.9	2787	1 TEL1_YEAST	P38110 saccharomyc
16	113	2.9	3744	1 TRAI_YEAST	P38811 saccharomyc
17	111.5	2.9	1332	1 IKAP_HUMAN	Q95163 homo sapien
18	111	2.9	607	1 DOR1_YEAST	Q04632 saccharomyc
19	111	2.9	1067	1 EG52_XENLA	Q91783 xenopus lae
20	110.5	2.9	2415	1 SPCA_MOUSE	P08032 mus musculu
21	109.5	2.8	564	1 M12_STRPY	P19401 streptococc
22	109.5	2.8	1066	1 KL61_DROME	P46863 drosophila
23	108.5	2.8	717	1 RH90_BRUPA	O61998 brugia paha
24	108.5	2.8	718	1 RHG8_HUMAN	Q61998 homo sapien
25	108.5	2.8	1442	1 DPO3_UREPA	Q9pqb4 ureaplasma
26	108.5	2.8	2672	1 GCN1_YEAST	P33892 saccharomyc
27	108	2.8	837	1 RA50_METTH	O26640 methanobact
28	108	2.8	1953	1 BN11_YEAST	P41832 saccharomyc
29	107.5	2.8	886	1 RA50_SULAC	Q33600 sulfolobus
30	107	2.8	1065	1 SEC8_YEAST	P32855 saccharomyc
31	106	2.8	900	1 IF38_ARATH	O49160 arabidopsis
32	106	2.8	925	1 Y562_HUMAN	O60308 homo sapien
33	106	2.8	1015	1 CND3_HUMAN	Q9bpx3 homo sapien

34	105.5	2.7	766	1 AMD3_MOUSE	O08739 mus musculu
35	105.5	2.7	1507	1 SIMA_DROME	Q24167 drosophila
36	105	2.7	732	1 HS9A_MOUSE	P07901 mus musculu
37	105	2.7	1312	1 RA50_YEAST	P12753 saccharomyc
38	105	2.7	8797	1 SNE1_HUMAN	Q8nf91 homo sapien
39	104.5	2.7	725	1 YK28_YEAST	P36114 saccharomyc
40	104.5	2.7	767	1 AMD3_HUMAN	Q01432 homo sapien
41	104.5	2.7	858	1 YNU1_YEAST	P40164 saccharomyc
42	104.5	2.7	1125	1 YE62_SCHPO	O14248 schizosacch
43	104.5	2.7	1134	1 A11A_HUMAN	P98196 homo sapien
44	104.5	2.7	1385	1 RRPO_PLAMV	Q07518 plantago as
45	104.5	2.7	4644	1 DYHC_MOUSE	Q9jhu4 mus musculu

ALIGNMENTS

RESULT 1  
MK10\_YEAST  
ID MK10\_YEAST STANDARD; PRT; 733 AA.  
AC Q02197;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glucose repressible protein MAK10.  
GN MAK10 OR YEL053C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93012936; PubMed=1398065;  
RA Lee Y.-J., Wickner R.B.;  
RT "MAK10, a glucose-repressible gene necessary for replication of a  
dsRNA virus of Saccharomyces cerevisiae, has T cell receptor  
alpha-subunit motifs.";  
RL Genetics 132:87-96(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=97313264; PubMed=9169868;  
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,  
Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";  
RL Nature 387:78-81(1997).  
CC -!- FUNCTION: Has a role in the propagation of L-A and M viruses,  
perhaps in the viral assembly. It is apparently directly needed  
for optimum respiration.  
CC -!- INDUCTION: Glucose-repressed.  
CC -!- SIMILARITY: SOME, TO T-CELL RECEPTOR ALPHA SUBUNIT VARIABLE  
REGIONS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M94533; AAA34749.1; --.  
EMBL; U18779; AAB64989.1; --.  
PIR; S31288; S31288.  
GermOnline; 139057; --.  
SGD; S0000779; MAK10.  
DR GO; GO:0004042; F:amino-acid N-acetyltransferase activity; IDA.  
DR GO; GO:0006474; P:N-terminal protein amino acid acetylation; IDA.



RN [11]  
RP VARIANT EL2 SER-28, AND VARIANT HPP ARG-48.  
RX MEDLINE=91346849; PubMed=1878597;  
RA Floyd P.B., Gallagher P.G., Valentino L.A., Davis M., Marchesi S.L.,  
RA Forget B.G.;  
RT "Heterogeneity of the molecular basis of hereditary  
RT pyropoikilocytosis and hereditary elliptocytosis associated with  
RT increased levels of the spectrin alpha I/74-kilodalton tryptic  
RT peptide.";  
RL Blood 78:1364-1372(1991).  
RN [12]  
RP VARIANT EL2 SER-45.  
RX MEDLINE=89323468; PubMed=2568862;  
RA Lecomte M.-C., Garbarz M., Grandchamp B., Feo C., Gautero H.,  
RA Devaux I., Bournier O., Galand C., D'Auriol L., Galibert F.,  
RA Sahr K.E., Forget B.G., Boivin P., Dhermy D.;  
RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white  
RT kindred with HE and HPP phenotypes.";  
RL Blood 74:1126-1133(1989).  
RN [13]  
RP VARIANT EL2/HPP PRO-207.  
RX MEDLINE=92176375; PubMed=1541680;  
RA Gallagher P.G., Tse W.T., Coetzer T., Lecomte M.-C., Garbarz M.,  
RA Zarkowsky H.S., Baruchel A., Ballas S.K., Dhermy D., Palek J.,  
RA Forget B.G.;  
RT "A common type of the spectrin alpha I 46-50a-kD peptide abnormality  
RT in hereditary elliptocytosis and pyropoikilocytosis is associated  
RT with a mutation distant from the proteolytic cleavage site. Evidence  
RT for the functional importance of the triple helical model of  
RT spectrin.";  
RL J. Clin. Invest. 89:892-898(1992).  
RN [14]  
RP VARIANT VAL-1857.  
RX MEDLINE=93253053; PubMed=8486776;  
RA Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,  
RA Kastally R., Kotula L., Delaunay J., Alloisio N.;  
RT "Low expression allele alpha LELY of red cell spectrin is associated  
RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and  
RT with partial skipping of exon 46.";  
RL J. Clin. Invest. 91:2091-2096(1993).  
RN [15]  
RP VARIANT EL2 BARCELONA PRO-469.  
RX MEDLINE=93372367; PubMed=8364215;  
RA dalla Venezia N., Alloisio N., Forissier A., Denoroy L., Aymerich M.,  
RA Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;  
RT "Elliptoikilocytosis associated with the alpha 469 His-->Pro  
RT mutation in spectrin Barcelona (alpha I/50-46b).";  
RL Blood 82:1661-1665(1993).  
RN [16]  
RP VARIANT CAGLIARI GLY-2024.  
RX MEDLINE=94043025; PubMed=8226774;  
RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,  
RA Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,  
RA Gallanello R.;  
RT "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta  
RT spectrin repeat 17 that severely disrupts the structure and self-  
RT association of the erythrocyte spectrin heterodimer.";  
RL J. Biol. Chem. 268:22656-22662(1993).  
RN [17]  
RP VARIANT EL2 CULOZ VAL-46, AND VARIANT EL2 LYON PHE-49.  
RX MEDLINE=90347052; PubMed=2384601;  
RA Morle L., Roux A.-F., Alloisio N., Pothier B., Starck J., Denoroy J.,  
RA Morle F., Rudigoz R.-C., Forget B.G., Delaunay J., Godet J.;  
RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I  
RT domain. Spectrin Culoz (GGT-->GTT; alpha I 40 Gly-->Val) and  
RT spectrin Lyon (CTT-->TTT; alpha I 43 Leu-->Phe).";  
RL J. Clin. Invest. 86:548-554(1990).  
RN [18]  
RP VARIANT EL2 JENDOUBA GLU-791.  
RX MEDLINE=92345619; PubMed=1638030;  
RA Alloisio N., Willmotte R., Morle L., Baklouti F., Marechal J.,  
RA Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,  
RA Delaunay J.;

RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is  
RT associated with elliptocytosis and carries a mutation distant from  
RT the dimer self-association site.";  
RL Blood 80:809-815(1992).  
RN [19]  
RP VARIANT EL2 TUNIS TRP-41.  
RX MEDLINE=89323436; PubMed=2568861;  
RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,  
RA Garbarz M., Dhermy D., Kastally R., Delaunay J.;  
RT "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due  
RT to the CGG-->TGG codon change (Arg-->Trp) at position 35 of the  
RT alpha I domain.";  
RL Blood 74:828-832(1989).  
RN [20]  
RP VARIANT EL2 GENOVA TRP-34.  
RX MEDLINE=94250920; PubMed=8193371;  
RA Perrotta S., del Giudice E.M., Alloisio N., Sciarratta G., Pinto L.,  
RA Delaunay J., Cutillo S., Iolascon A.;  
RT "Mild elliptocytosis associated with the alpha 34 Arg-->Trp mutation  
RT in spectrin Genova (alpha I/74).";  
RL Blood 83:3346-3349(1994).  
RN [21]  
RP VARIANT EL2 ANASTASIA THR-45.  
RX MEDLINE=95290423; PubMed=7772539;  
RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,  
RA Cutillo S., del Giudice E.M.;  
RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45  
RT Arg-->Thr) with moderate elliptocytogenic potential.";  
RL Br. J. Haematol. 89:933-936(1995).  
CC -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal  
CC network underlying the erythrocyte plasma membrane. It associates  
CC with band 4.1 and actin to form the cytoskeletal superstructure of  
CC the erythrocyte plasma membrane.  
CC -!- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which  
CC aggregate side-to-side in an antiparallel fashion to form dimers,  
CC tetramers, and higher polymers.  
CC -!- DISEASE: Defects in SPTAL are a cause of rhesus-unlinked  
CC elliptocytosis 2 (EL2) [MIM:130600, 182860, 166900].  
CC Elliptocytosis (also known as ovalocytosis) is a genetically  
CC heterogeneous, autosomal dominant hematologic disorder. It is

Query Match 3.5%; Score 134; DB 1; Length 2418;  
Best Local Similarity 19.4%; Pred. No. 0.21;  
Matches 175; Conservative 142; Mismatches 307; Indels 280; Gaps 48;  
QY 13 GWELSMPEKMEKSNTNW---VDITQD----FEEACRELKLGELLHDKLFGLFEEAMSAIEM 65  
Db 1572 GNEEAMKEQLKEHWDHLLERTNDKGKLNESARQQRNTSIRDFEFLWSEAEITLAM 1631  
QY 66 MDPKMDAGMIGN-----QVNRKVLNFEQAIKDGTIKIKOLTPELIGIMDTFCCLITW 119  
Db 1632 KDQARDLASAGNLLKKHQLLEREMLAAREDAKDLNTLAE DL----- 1672  
QY 120 LEGHSLAQTVFTCLYIHNPDFI---EDPAMKAFALGILKICDIAREKVNKA-AVF----- 170  
Db 1673 -----LSSGTF-----NVDQIVKKKDNVKNRF-LNVQELAAAHHEKLKEAYALFOFFQD 1720  
QY 171 --EEEDF-----QSMTYGFKMANSVTDL-----RVTGMLKDVED-----DMQRRVK 209  
Db 1721 LDDEESWTEEKLIIRVSSQDYG-RDLQGVQNLLKKHKRLEGELVAHEPAIQNVLDMAEKLK 1779  
QY 210 STRSRQGERDPPEVEL-----EHQQCLAVFSRVKFTRVLLTV-----LIAFTKKETSAVA 259  
Db 1780 D-KAAVGQE--EIQRLAQFVEHWEKLEKARGLKLEESLEYLQFMQNAEEEEAWIN 1835  
QY 260 EAQKLMVQ--AADLLSAI-----HNSLH-----HGIQAQNDTTKGDHPIMMGFEPLVNQ 306  
Db 1836 EKNALAVRGDCGDTLAATQSLMKHEALENDFAVHETRVQNVCAQG-----EDILNK 1887  
QY 307 RLLPPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCLILDF-FCEESEQSPCV 365  
Db 1888 VLHEES-----QNKEISSKIEALNEKTPSLAKAIAAWKQLLEDYAFQEFNWKADV 1939



QY 366 LSRSLQTTFLVDNKKVFGTHL-----MQDMVKDALRSFVD---PPVLSPKCVLYN- 413  
Db 1940 EAWIADKETSLSKTNGN--GABLDGFLTLAKQDTLDASLSQSFQOERLPEITDLKDKLISA 1997  
QY 414 NHQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQ---DEFMTFFYF- 469  
Db 1998 QHNSKAIEERYAALLKRWEQLEASAVHRQK-----LLEKQLPLQKAEDLFVEFAHK 2050  
QY 470 -----NRAEKVDAAL---HTMLLKQ--EPQRQH-----LAC 495  
Db 2051 ASALNNWCEKMEENLSEPVHCVSLNEIRQLQKDHFDFLASLARAQADFKCLLELDQIKA 2110  
QY 496 LG-----TWVLYHNLIRIMIQYL-----LSGFEL-ELYSMHEYIYIY 530  
Db 2111 LGVPSSPYTWLTVEVLRTWKHLSDIIEEREQELQKEEARQVKNFEMCQEFQONASTFLQ 2170  
QY 531 WYLSEFLYAWLMSTLSRADGSQMABERIMEEQKGRSSKKTKKKKVRPLSREIT----- 585  
Db 2171 WILETRAYF-----LDGSLKETGTLESQLEANK-----RKQKEIQAMKRQLTKIVDL 2218  
QY 586 -----MSQAYQ---NMCAQMFXTMVAFDMDGKVRKPKFELDSE 620  
Db 2219 GDNLEDALILDIKYSTIGLAQWDLQYQLGLRMQHNLQEQIQAKDIKGVSEETLKEPST- 2277  
QY 621 QVRYEHRFAPFNSVMTPPPVHYLQFKE-MSDLNKYSPPPQSPEL----- 663  
Db 2278 --IYKH---FDENLTGRLTH-KEFRSCLRGLNYLPMVEEDEHEPEKFEKFLDAVDPRGK 2330  
QY 664 -YVAASKHFQAKMI---LENIPNDHEVNRILKVAKPNFVVMKLLAGGHKKESKVPPEF 719  
Db 2331 GYVSLEDY--TAFLLDKESENIKSSDEIENAFQALAEKGSYITK---EDMKQALTPEQV 2384  
QY 720 DESA 723  
Db 2385 SFCA 2388

RESULT 3

SAB2\_HUMAN STANDARD; PRT; 733 AA.  
AC Q9UPW6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE DNA-binding protein SATB2 (Special AT-rich sequence-binding protein  
DE 2).  
GN SATB2 OR KIAA1034.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
RN [2]  
RP DISEASE.  
RX MEDLINE=22855719; PubMed=12915443;  
RA FitzPatrick D.R., Carr I.M., McLaren L., Leek J.P., Wightman P.,  
RA Williamson K., Gautier P., McGill N., Hayward C., Firth H.,  
RA Markham A.F., Fantes J.A., Bonthron D.T.;  
RT "Identification of SATB2 as the cleft palate gene on 2q32-q33.";  
RL Hum. Mol. Genet. 12:2491-2501(2003).  
CC -!- FUNCTION: May play an important role in palate formation.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- TISSUE SPECIFICITY: High expression in adult brain, moderate  
CC expression in fetal brain, and weak expression in adult liver,

CC kidney, and spinal cord and in select brain regions, including  
CC amygdala, corpus callosum, caudate nucleus, and hippocampus.  
CC -!- DISEASE: Involved in isolated cleft palate through 2 chromosomal  
CC translocations t(2;7) and t(2;11).  
CC -!- SIMILARITY: Belongs to the CUT homeobox family.  
CC -!- SIMILARITY: Contains 2 CUT domains.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB028957; BAA82986.1; ALT\_INIT.  
DR Genew; HGNC:21637; SATB2.  
DR MIM; 608148; -.  
DR MIM; 119540; -.  
DR InterPro; IPR007108; Cut\_homeo.  
DR InterPro; IPR003350; Hmoeo CUT.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF02376; CUT; 2.  
DR Pfam; PF00046; homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; FALSE\_NEG.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Nuclear protein; Repeat;  
KW Chromosomal translocation.  
FT DNA\_BIND 350 437 CUT 1.  
FT DNA\_BIND 473 560 CUT 2.  
FT DNA\_BIND 615 674 HOMEBOX.  
SQ SEQUENCE 733 AA; 82555 MW; 1FE1FCBD34F11E9E CRC64;

Query Match 3.3%; Score 127; DB 1; Length 733;  
Best Local Similarity 20.9%; Pred. No. 0.14;  
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;  
QY 217 EERDPEVEL---EHQQCLAVFSRVKFTRVLLTVLIAFTKETSAAVAEQKLM----- 265  
Db 67 EQLDGSLEYDNREEHAEFVLRKDVLFSLQVETALLALGYSHSSA-AQAQGIKLGWNP 125  
QY 266 -----VQAADLLSAIHNSLHHGIIQAQNDTTKGHP-----IMMGFEPL--- 303  
Db 126 LPLSYVTDAPDATVADMLQDVYHVVTLKIQLOSCSKLEDLPAEQWNHATVRNALKELLE 185  
QY 304 VNQRLLPPTFPRIYAKIIRHEEMVNYFA-----RLIDRIKTV-CEVNLTLN--H 349  
Db 186 MNQSTLAKECPLSQSMISSIVNSTYANVSATKCKQEFGRWYKYYKKIKIVERENLSY 245  
QY 350 CILDFFCFESEQSPCVLSR-----SLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPV 404  
Db 246 CVLG-----QRPMLHPNMNQLASLGKTNEQSPHSQIHSTPIRNQV-PALQPMSPGL 297  
QY 405 LSPK-----CYLYNN-----HQAKDCID-SFVTH-----CVR--PFCSL 435  
Db 298 LSPQLSPQLVRQQAIAHNLNQIAVSRLLAHQHPQAINQQFLNHPPIPRAVKPEPTNSS 357  
QY 436 IQIHGHNRRARQDKLGHILEEFATLQDEFMTFFYENRAEKVDAALHTMLLKQEPQRQLAC 495  
Db 358 VEVSPDIYQQVRDE---LKRASVSQAVFARVAFNRQTQ---GLLSEILRKEEDPR----- 405  
QY 496 LGTWVLYHNLIRIMIQYLLSGFELE---LY-----SMHEYIYIYWLSE----- 535  
Db 406 TASQSLLVNLRAM-QNFLNLPVEVERDRIYQDERERSMNPNVSMVSSASSSSSRTPOAK 464  
QY 536 -----FLYAWLMSTLSRADGSQMABERIMEEQKGRSSKKTKKK 574  
Db 465 TSTPTDLPKVDGANINITAAIYDEIQEMKRAKVSQALFAKVAANKSQGLCELLRWK 524  
QY 575 KKVRLPSREITMSQAYQNMCAGMFKTMVAFDMDGKVRKPKFELD---SEQVRYEHRFAPF 631

Db 525 ENPSPENRTL-----WENLC-----TIRRF-----LNLPOHERDVIYEEESRHHHSERMQ 569

QY 632 NSVMTpp-pVHYLQ-----FKEMSDLNKYSPPPPQSPELYVAASKHFQQAkmILENIP-- 682

Db 570 HVVQLPPEPVQVLHRQSQPAKESPPPEEAPPPPTEDSCAKKPRSRtkISLEALGIL 629

QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703

Db 630 QSFHVDGLYPDQEAHHTLSAQLDLPKHTIIK 661

RESULT 4

SAB2\_MOUSE

ID SAB2\_MOUSE STANDARD; PRT; 733 AA.

AC Q8VI24;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE DNA-binding protein SATB2 (Special AT-rich sequence-binding protein

DE 2).

GN SATB2 OR KIAA1034.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.

RX MEDLINE=22855719; PubMed=12915443;

RA FitzPatrick D.R., Carr I.M., McLaren L., Leek J.P., Wightman P.,

RA Williamson K., Gautier P., McGill N., Hayward C., Firth H.,

RA Markham A.F., Fantes J.A., Bonthron D.T.;

RT "Identification of SATB2 as the cleft palate gene on 2q32-q33.";

RL Hum. Mol. Genet. 12:2491-2501(2003).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Embryonic tail;

RA Okazaki N., Kikuno R., Inamoto S., Koseki H., Hiraoka S.,

RA Saga Y., Nagase T., Ohara O., Koga H.;

RT "Prediction of the coding sequences of mouse homologues of KIAA gene:

RT III. The complete nucleotide sequences of 500 mouse KIAA-homologous

RT cDNAs identified by screening of terminal sequences of cDNA clones

RT randomly sampled from size-fractionated libraries.";

RL DNA Res. 10:167-180(2003).

CC -!- FUNCTION: May play an important role in palate formation.

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q8VI24-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q8VI24-2; Sequence=VSP\_008967;

CC Note=No experimental confirmation available;

CC -!- DEVELOPMENTAL STAGE: Expression first detected at 10.5 dpc in the

CC maxillary component of the first pharyngeal arch and the lateral

CC aspect of the frontonasal process in the regions that will

CC subsequently fuse to form the primary palate. At 11 - 11.5 dpc,

CC the expression pattern demarcates the region of the medial aspect

CC of the maxillary process within the primitive oral cavity, which

CC will form the palate shelf. By 12.5 dpc, symmetrical expression is

CC seen in the medial edges of the developing palate shelves and this

CC continues until 13.5 dpc when the strongest expression is in the

CC mesenchyme underlying the medial edge epithelia. By the time of

CC palatal shelf fusion at 14.5 dpc the expression is dramatically

CC down-regulated. No expression detected elsewhere in the embryo at

CC any stage examined.

CC -!- SIMILARITY: Belongs to the CUT homeobox family.

CC -!- SIMILARITY: Contains 2 CUT domains.

CC -!- SIMILARITY: Contains 1 homeobox domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF319623; AAL37172.1; --

DR EMBL; AK129270; BAC98080.1; ALT\_INIT.

DR InterPro; IPR007108; Cut\_homeo.

DR InterPro; IPR003350; Hmoec\_CUT.

DR InterPro; IPR001356; Homeobox.

DR Pfam; PF02376; CUT; 2.

DR Pfam; PF00046; homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.

DR PROSITE; PS50071; HOMEBOX\_2; 1.

KW Developmental protein; Homeobox; DNA-binding; Nuclear protein; Repeat;

KW Alternative splicing.

FT DNA\_BIND 350 437 CUT 1.

FT DNA\_BIND 473 560 CUT 2.

FT DNA\_BIND 615 674 HOMEBOX.

FT VARSPPLIC 58 116 Missing (in isoform 2).

FT /FTid=VSP\_008967.

SQ SEQUENCE 733 AA; 82559 MW; 153CFD1CC3491F25 CRC64;

Query Match 3.3%; Score 127; DB 1; Length 733;

Best Local Similarity 20.9%; Pred. No. 0.14;

Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

QY 217 EERDPEVEL-----EHQQCLAVFSRVKFTRVLLTVLIAFTKETSAAVAEAQKLM----- 265

Db 67 EQLDGSLEYDNREEHAEFVLVRKDVLFSQLVETALLALGYSHSSA-AQAQGIKLRWNP 125

QY 266 -----VQAADLLSAIHNSLHHGIQAQNDTTKGDHP-----IMMGFEPL--- 303

Db 126 LPLSYVTDAPDATVADMLQDVYHVVTLKIQQLQCSKLEDLPAEOWNHATVRNALKELLKE 185

QY 304 VNQRLLPPTFPYAKIIKREEMVNYEA-----RLIDRIKTV-CEVNLTNL--H 349

Db 186 MNQSTLAKECPLSQSMISSIVNSTYIANVSATKCEFGRWYKKYKIKIVERVERENLSY 245

QY 350 CILDFCFEFSQSPCVLSR-----SLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPV 404

Db 246 CVLG-----QRPMHLPMNMQLASLGKTNEQSPHSQIIHHSHTPIRNQV-PALQIPMS 297

QY 405 LSPK-----CYLYNN-----HQAKD CID-SFVTH-----CVR--PFCSL 435

Db 298 LSPQLSPQLVRRQIIAMAHLINQIIAVSRLLAHQHPQAINQQFLNHPPIPRAVKPEPTNSS 357

QY 436 IQIHGHNRAQRDKLGHILEEFATLQDEFMTFYFNRAEKVDAALHTMLLKQEPQRQHLAC 495

Db 358 VEVSPDIYQQVRDE----LKRASVSQAVFARVAFNRTQ---GLLSEILRKEEDPR----- 405

QY 496 LGTWVLYHNLIRIMIYLLSGFELE---LY-----SMHEYYYIYWYLSE----- 535

Db 406 TASQSLLVNLRAM-QNFLNLPEVERDRYQDERERSMNPVSMVSSASSPSSSRTPQAK 464

QY 536 -----FLYAWLMSTLSRADGSQMAEERIMEEQQKGRSSKTKKK 574

Db 465 TSTPTTDLPIKVDGANVNITAAIYDEIQQEMKRAKVSQALFAKVAANKSQGLCELLRWK 524

QY 575 KKVRLPSREITMSQAYQNMCAGMFKTMVAFDMGKVRKPKFELD---SEQVRYEHRFAPF 631

Db 525 ENPSPENRTL-----WENLC-----TIRRF-----LNLPOHERDVIYEEESRHHHSERMQ 569

QY 632 NSVMTpp-pVHYLQ-----FKEMSDLNKYSPPPPQSPELYVAASKHFQQAkmILENIP-- 682

Db 570 HVVQLPPEPVQVLHRQSQPTKESPPPEEAPPPPTEDSCAKKPRSRtkISLEALGIL 629

QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703

Db 630 QSFHVDGLYPDQEAHHTLSAQLDLPKHTIIK 661

RESULT 5

NMD2\_YEAST  
ID NMD2\_YEAST STANDARD; PRT; 1089 AA.  
AC P38798;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nonsense-mediated mRNA decay protein 2 (Up-frameshift suppressor 2).  
GN NMD2 OR UPE2 OR IFS1 OR SUAI OR YHR077C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95189083; PubMed=7883168;  
RA He F., Jacobson A.;  
RT "Identification of a novel component of the nonsense-mediated mRNA  
decay pathway by use of an interacting protein screen.";  
RL Genes Dev. 9:437-454 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=PLY136;  
RX MEDLINE=95189082; PubMed=7883167;  
RA Cui Y., Hagan K.W., Zhang S., Peltz S.W.;  
RT "Identification and characterization of genes that are required for  
the accelerated degradation of mRNAs containing a premature  
translational termination codon.";  
RL Genes Dev. 9:423-436 (1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95327692; PubMed=7604038;  
RA Lee S.I., Umen J.G., Varmus H.B.;  
RT "A genetic screen identifies cellular factors involved in retroviral  
-1 frameshifting.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:6587-6591 (1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288c / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,  
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,  
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
VIII.";  
RL Science 265:2077-2082 (1994).  
CC -!- FUNCTION: Involved in nonsense-mediated decay of mRNAs containing  
premature stop codons. It interacts, via its C-terminus, with  
NAM7/UPF1. Could be involved in determining the efficiency of  
translational termination or reinitiation or factors involved in  
the initial assembly of an initiation- and termination-competent  
mRNP.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; U14974; AAA67724.1; -.  
CC EMBL; U12137; AAA66521.1; -.  
CC EMBL; U28158; AAA74948.1; -.  
CC EMBL; U10556; AAB68893.1; -.  
CC PIR; S48244; S48244.  
CC GermOnline; 139394; -.  
CC SGD; S000119; NMD2.  
CC GO; GO:0005844; C:polysome; IDA.

DR InterPro; IPR008938; ARM.  
DR InterPro; IPR003890; IF\_eIF4G.  
DR InterPro; IPR007193; Upf2.  
DR Pfam; PF02854; MIF4G; 3.  
DR Pfam; PF04050; Upf2; 1.  
DR SMART; SM00543; MIF4G; 3.  
FT DOMAIN 843 975 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT CONFLICT 2 2 D -> YQQ (IN REF. 3 AND 4).  
SQ SEQUENCE 1089 AA; 126746 MW; 13BBE725675CBF52 CRC64;  
  
Query Match 3.3%; Score 127; DB 1; Length 1089;  
Best Local Similarity 17.6%; Pred. No. 0.24;  
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;  
  
QY 24 KSNTNWV-----DITQDFEEACRELKLGELL-----HDKLFGLF 57  
| | | | | : : : : : | | | | | : : : : :  
Db 35 KRNTGFIKKLKKGVKGSSESLKDLSEASLEKYLSEIIVTVTECLLNVLNKNDDVIAAV 94  
| | | | | : : : : : | | | | | : : : : :  
QY 58 EAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIKDGIKIKDLTLPELIGIMDTCFCCCL 116  
| | | | | : : : : : | | | | | : : : : :  
Db 95 EIIISGLHQRFNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR 136  
| | | | | : : : : : | | | | | : : : : :  
QY 117 ITWLEGHSLAQTVFTCLYIHN-----PDFIE-----DPAMKAFALGILK- 155  
| | | | | : : : : : | | | | | : : : : :  
Db 137 ITRVKGN---LRVFTELYLVGVFTLDDIESKDAIPNLFQKTKGRKDPLLFSILREILNY 193  
| | | | | : : : : : | | | | | : : : : :  
QY 156 -----ICDIAREKVNK-AAVFEED--FQSMTYGFKMANSVTDL-----RVTGM 196  
| | | | | : : : : : | | | | | : : : : :  
Db 194 KFKLGFTTTIATAFIKKEAPLFRDDDDNSWDDLIYDSKLGALQSLFKNFIDATFARATEL 253  
| | | | | : : : : : | | | | | : : : : :  
QY 197 LKDVEDDMQRRVVKSTRSQGEERDPEVELEHQCLAVFSRVKFTRVLLTVLIAFTKXETS 256  
| | | | | : : : : : | | | | | : : : : :  
Db 254 HKKV-NKLOREHQKCCQITGKLREYVE-EYDKLLEFIRFK-----TS 295  
| | | | | : : : : : | | | | | : : : : :  
QY 257 AVAEAQKLMVQAADLLSAIHNSLHHGIAQNDDTTKGHPIMMGFEPLVNQRLLPPTPPRY 316  
| | | | | : : : : : | | | | | : : : : :  
Db 296 AITLGEFFKLEIPELEGASNDL-----KETA-----SPMITNQILLPPN--- 334  
| | | | | : : : : : | | | | | : : : : :  
QY 317 AKIKREEMVNYFARLIDRIKTVCEV-----VNLTNLHCILDFCFEFSQSPCVL 366  
| | | | | : : : : : | | | | | : : : : :  
Db 335 QRLWENEDTRKFEYILPDISKTVESQSSTKEDSNVNSKNINL---FFTDL-EMADC-- 388  
| | | | | : : : : : | | | | | : : : : :  
QY 367 SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA 417  
| | | | | : : : : : | | | | | : : : : :  
Db 389 -----XDIIDLSNRYWSSYLDNKAIRNLKFFMETQDWSKLPVYS-RFIATNSKYM 440  
| | | | | : : : : : | | | | | : : : : :  
QY 418 KDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFMTFFYNRAEKVDA 477  
| | | | | : : : : : | | | | | : : : : :  
Db 441 PEIVSEFINYLDNGFRS--QLHS-----NKNVKNVI 469  
| | | | | : : : : : | | | | | : : : : :  
QY 478 ALHTMLLKQEPQORQHLAGLGTWVLYHNLIRIMIQYLLSGFELEYSMHEYYIYWYLSSEFL 537  
| | | | | : : : : : | | | | | : : : : :  
Db 470 IFFSEMIK-----FQLIPSEFMIFHKIRTLIMYQVNNVEILTV-----LLEHSGKFL 517  
| | | | | : : : : : | | | | | : : : : :  
QY 538 YAWLMSTLSRADGSQMAERIMEEQKGRSSKTKKKKKVVRPLSREITMSQAYQNMCMAGM 597  
| | | | | : : : : : | | | | | : : : : :  
Db 518 -----LNKPEYKELMEKMW-----QLIKDKKNDROLN--MNMKSALENITLL 558  
| | | | | : : : : : | | | | | : : : : :  
QY 598 FKTMY 602  
| | | | | : : : : : | | | | | : : : : :  
Db 559 YPPSV 563  
| | | | | : : : : : | | | | | : : : : :  
  
RESULT 6  
VIT3\_CAEEEL STANDARD; PRT; 1603 AA.  
ID VIT3\_CAEEEL  
AC Q9N4J2;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Vitellogenin 3 precursor.  
GN VIT-3 OR F59D8.1.  
OS Caenorhabditis elegans.











RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX MEDLINE=89210800; PubMed=3072195;  
RA Lemaire C., Heilig R., Mandel J.L.;  
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal  
RT coding and 3' untranslated regions between man and chicken.";  
RL EMBO J. 7:4157-4162(1988).  
CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the  
CC plasma membrane.  
CC  
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY  
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,  
CC ABP-120, ABP-180, OR BETA-FODRIN).  
CC  
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC  
CC -!- SIMILARITY: Contains 22 spectrin repeats.  
CC  
CC -!- SIMILARITY: Contains 1 WW domain.  
CC  
CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X13369; CAA31746.1; -.  
CC PIR; S02041; S02041.  
CC HSSP; P46939; 1BHD.  
CC InterPro; IPR001589; Actbind actnin.  
CC InterPro; IPR001715; Calponin-like.  
CC InterPro; IPR002017; Spectrin.  
CC InterPro; IPR001202; WW\_Rsp5\_WWP.  
CC InterPro; IPR000433; Znf\_ZZ.  
CC Pfam; PF00307; CH; 2.  
CC Pfam; PF00435; spectrin; 21.  
CC Pfam; PF00397; WW; 1.  
CC Pfam; PF00569; ZZ; 1.  
CC SMART; SM00033; CH; 2.  
CC SMART; SM00150; SPEC; 21.  
CC SMART; SM00456; WW; 1.  
CC SMART; SM00291; Znf\_ZZ; 1.  
CC PROSITE; PS00019; ACTININ\_1; 1.  
CC PROSITE; PS00020; ACTININ\_2; 1.  
CC PROSITE; PS50021; CH; 2.  
CC PROSITE; PS01159; WW\_DOMAIN\_1; 1.  
CC PROSITE; PS50020; WW\_DOMAIN\_2; 1.  
CC PROSITE; PS01357; ZF\_ZZ\_1; 1.  
CC PROSITE; PS50135; ZF\_ZZ\_2; 1.  
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;  
KW Repeat; Zinc-finger.  
FT DOMAIN 1 244 ACTIN-BINDING.  
FT DOMAIN 19 123 CH 1.  
FT DOMAIN 138 241 CH 2.  
FT REPEAT 341 449 SPECTRIN 1.  
FT REPEAT 450 558 SPECTRIN 2.  
FT REPEAT 561 669 SPECTRIN 3.  
FT REPEAT 721 830 SPECTRIN 4.  
FT REPEAT 832 936 SPECTRIN 5.  
FT REPEAT 945 1047 SPECTRIN 6.  
FT REPEAT 1050 1156 SPECTRIN 7.  
FT REPEAT 1159 1265 SPECTRIN 8.  
FT REPEAT 1268 1369 SPECTRIN 9.  
FT REPEAT 1470 1570 SPECTRIN 10.  
FT REPEAT 1573 1678 SPECTRIN 11.  
FT REPEAT 1681 1782 SPECTRIN 12.  
FT REPEAT 1879 1981 SPECTRIN 13.  
FT REPEAT 2013 2103 SPECTRIN 14.  
FT REPEAT 2106 2211 SPECTRIN 15.  
FT REPEAT 2214 2321 SPECTRIN 16.  
FT REPEAT 2472 2574 SPECTRIN 17.  
FT REPEAT 2577 2683 SPECTRIN 18.  
FT REPEAT 2686 2799 SPECTRIN 19.  
FT REPEAT 2802 2904 SPECTRIN 20.

FT REPEAT 2906 2928 SPECTRIN 21.  
FT REPEAT 2931 3037 SPECTRIN 22.  
FT DOMAIN 3052 3085 WW.  
FT ZN FING 3304 3351 ZZ-TYPE.  
FT VARIANT 1171 1171 MISSING.  
FT VARIANT 1869 1869 Q -> H.  
FT VARIANT 1885 1885 K -> R.  
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;  
  
Query Match 3.1%; Score 117.5; DB 1; Length 3660;  
Best Local Similarity 17.7%; Pred. No. 6;  
Matches 132; Conservative 113; Mismatches 229; Indels 273; Gaps 35;  
  
QY 13 GWELSMPEKMEKSNNTNWVDIT-----  
Db 1646 GKESLVEDKLSLLNSNWIATVSRAEELNLLMEYQKHMEAFDQKVANVTTWIYRAEILLD 1705  
  
QY 34 -----QDFEEACRELKLGEL--LHDKLFGLFEEAMSAIEMM-----DPKMDAG 73  
Db 1706 ESDKQKPQKEETLKRK-AELNDMPKVDSDVRD--QAVDLMTNRGDHCRKVIEPKL--- 1759  
  
QY 74 MIGNQVNRKVLNFEQAIKDGTIKIKDLTLPELIGIMDTFCCLITWLEGHSLAQTVFTCL 133  
Db 1760 ---SELNHRFAAISQRIKSG-----  
  
QY 134 YIHPDFIEDPAMKAFALGILKICDIAREKYNKAAVFEEDF-----QSMTY 180  
Db 1777 ----KPFIPLKELEQDFDIQKLLLEPLEVEIQGVNLFKEEDFNKDMSEDDSTVKELLQR 1832  
  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDPEVELEHQQLAVFSRVKFT 240  
Db 1833 GDTLQKRITDERKREEIKIKQQLLTQKHNALKDLSQRRKKALEISHQW----YQYKRQA 1888  
  
QY 241 RVLLTVLIAFTKETSAV--AEAQKLMVQAADL-----LSAIHNSLHHGIAQNDDTKG 292  
Db 1889 DDLMTWLDLIEKKLASLPDHKDEQKLKEIGGELEKKKEDLNANR-----QAERLSKDG 1942  
  
QY 293 DHPIMMGFEPLVNQRLPPTFPFYAKIIKREEMVNVYFARLIDRIKTVCEVVNLTNLHCIL 352  
Db 1943 -----AAKAVEPTLVQLSK--RWRDFESKFAQF-----RRLNVAQIQTVL 1980  
  
QY 353 DFFCEFESEQPCVLSRL-LQTTFLVDNKKVFGTHLMQDM--VKDALRSFVDPVLPVSPKC 409  
Db 1981 -----EDTTFVMTESMTVETTVVPSTYLAELQLLQALSEVEERLNS---PVL---- 2025  
  
QY 410 YLYNNHOAKDCIDSF-VTHCVRPF--CSLIQIHGHNRRARQDKLGHIL-----EEFATLQD 462  
Db 2026 -----QAKDCEDLLKQEECLKNIKDC-LGRLOQHIDIIHKKKTPALQSATPRETANIQD 2078  
  
QY 463 EFMTFYFNRAEKVDAAALHTMLLKQEPQORQHACLGTVVLYHNLRIMIQYLLSGFELELYS 522  
Db 2079 K-LTQLNSQWEKVNMYRDRQARFDKSKE-----KWRLFH-----CEMKS 2117  
  
QY 523 MHEYYIYWYLSEFLYAWLMST---LSRA-----DGSQMAEERIMEEQKGRSSKTKKKK 574  
Db 2118 FNE-----WLTETEEKLSRAQIEAGDVGHVKTQKFLQELQDGIGRQQTVVVK 2163  
  
QY 575 KKVRLPSREITMSQAYQNMCMAGMFKTMVAFDMGKVRKPKF-----ELDSEQVRYEH 626  
Db 2164 -----TLNVTGEEIIEQSSAA-----DANVLKEQLGNLNRWQEICRQLVEKRRK 2207  
  
QY 627 RFAPFNSVMTPPPVHYLQFKEMSDLNK 653  
Db 2208 RIEEKNILS-----EFQE--DLNK 2225  
  
RESULT 11  
BAC1\_HUMAN  
ID BAC1\_HUMAN STANDARD; PRT; 736 AA.  
AC O14867; O43285;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)







Db 718 YEIS---XKKETASFIPTKESLIRDFEQCCCKEKELQMLRKESEISHN----- 763

Qy 63 IEMDPKMDAGMIGNQVRKVL-----NFEQAIKDGTIKIKDLTLPELIGIMDTFCCLIT 118

Db 764 ----ENKMDPSSKEGQYKAKIKELENNLERLSDLQSKIQE-----IESIRCKDSQLK 813

Qy 119 WLEGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFE----- 171

Db 814 WAQ-----NTIDDTMKMKSL-----LTLSNKETTIEKLSSEIE 848

Qy 172 --BEDFQSMTYGFKMANSVTDL-----RVTGMLKDVEDDMQR----- 206

Db 849 NLDKELRKTKFOYKFLDQNSDASTLEPTLRKELEQIQVQLKDANSQIQAYEEIISSNENA 908

Qy 207 --RVKSTRSRQGEERDPEVELEHQOCLA----- 232

Db 909 LIELKNELAKTKENYDAKIELEKKEKWAREEDLSRLRGELGEIRALQPKLKEGALHFVQQ 968

Qy 233 -----VFSRVKFTRVLLTVLIAFTTKKETSAAVEAQKLMVQAADLLSAIHNSLHHG 282

Db 969 SEKLNEVERIQMKIEKMKSTIVQLCKKKEMS---QYQSTMKENKDLSELVIRLEKDA 1025

Qy 283 IQAQNDTTKGDPIMMGFEPLVNPQLPPTFPFYAKIIKREEMVNYFARLIDRIKTV--C 340

Db 1026 ADCQALTKTKSSLYSA-----QDLLDKHERKW-----MEEKADYERELISNIEQTESL 1074

Qy 341 EVNLTNLHCILDDFFCEFSQSPCVL-----SRSLQTTFVLVDNKKVFGTHLMQDM 391

Db 1075 RVENSVLTEKVDVDDTAANNQDKHLKLVSLFSLNRHERNSLETKLTCTCKRELAFVKQKND 1134

Qy 392 VKDALRSFVDPVPLSPKCYLYNNHQAKDC---IDSF--VTHCVRPFCSLIQIHGHNRR 445

Db 1135 LEKTINDLQRTQTLSKEY-----QCSAVIIDEFKDITKEV----TQVNIKENNAI 1182

Qy 446 QRDKLGHILEE----FATLQDEFMTFYNRAEKVDAALHTMLLKQEPQROHLACLGTWVL 501

Db 1183 LQKSLKNVTEKNREIYKQLND-----RQEII-SRLQRDLIQTKEQ-----VS 1223

Qy 502 YHNLRLIMIQYLLSGFELELYSMHEYYY-----IYWVLSSEFLYAWLMSTLSRA 548

Db 1224 INSNKILV-----YESEMEQCKQRYQDLSQQQKDAKKDIEKLTNEI--SDLKGLSSA 1275

Qy 549 DGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNNMCAG--MFKTMVAFDM 606

Db 1276 ENANADLENKFNRLKKQAHEKLDASKQQAALTNELNELKAIDKLEQDLHFENAKVIDL 1335

Qy 607 DGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPVHYLQFKEMSDLNKYSPPPQSPELYV- 665

Db 1336 DTKLK--AHBLQSEDVSRDHEKDTYRTLK-----EEIESLKK-----ELQIF 1375

Qy 666 ----AASKHFQQAKMILENIPNDHEVNRIL 692

Db 1376 KTANSSSDAFPEKLV-----NMEKEKDRII 1400

RESULT 14

SPOF\_SCHPO

ID\_SPOF\_SCHPO STANDARD; PRT; 1957 AA.

AC Q104I1; Q9USE9;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sporulation-specific protein 15.

GN SPO15 OR SPAC1F3.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.

RX MEDLINE=20107136; PubMed=10639340;

RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;

RT "S. pombe sporulation-specific coiled-coil protein Spol5p is localized to the spindle pole body and essential for its modification.";

RI J. Cell Sci. 113:545-554 (2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880 (2002).

RN [3]

RP SEQUENCE OF 705-871 FROM N.A.

RC STRAIN=968 h90;

RX MEDLINE=20223868; PubMed=10759889;

RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T., Hiraoka Y.;

RT "Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library.";

RI Genes Cells 5:169-190 (2000).

CC -!- FUNCTION: Has a role in the initiation of spore membrane formation.

CC -!- SUBUNIT: Monomer.

CC -!- SUBCELLULAR LOCATION: Spindle pole body.

CC -!- SIMILARITY: Belongs to the MPC70 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z70690; CAA94624.1; --.

DR EMBL; AB027811; BAA87115.1; --.

DR PIR; T38077; T38077.

DR GenedB\_SPombe; SPAC1F3.06C; --.

KW Sporulation; Coiled coil.

FT DOMAIN 199 785 COILED COIL (POTENTIAL).

FT DOMAIN 804 1235 COILED COIL (POTENTIAL).

FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).

FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).

SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 2.9%; Score 113.5; DB 1; Length 1957;

Best Local Similarity 21.1%; Pred. No. 5.1;

Matches 113; Conservative 86; Mismatches 215; Indels 121; Gaps 24;

Qy 9 DDDSGWELSMPEKMSNTNWDITQDFEACRELKLGEL--LHDKLFGLFAMSAIEMM 66

Db 1449 DDTSRKNSSLMKIESINSSLDD--KSFELASAVEKLGALQKHLHSESLSMENIKS-QLQ 1505

QY 67 DPK-----MDAGMI---GNQVNRKVLNFEQAIKDGTIKIKDLT-----LPELIGIMDTCFEC 114

Db 1506 EAKEKIQVDESTIQELDHEITASKNNYEGKLNKDKSIRDLSENIEQLNNLLAEKSAVK 1565

QY 115 CLITWLEGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAPFEE-- 172

Db 1566 RLSTEKESAILQFNRLADLEYHKSQVES-----ELGRSKLKLASTTEELQ 1611

QY 173 ----EDFQSMTYGFKMANSVTDLRVTGMLKOVEDMQRRVKSTRSRQGEERDPEVELEHQ 228

Db 1612 LAENERLSLTTTMLDLQNVKD-----LSNIKDSLSEDLRTLRSLE----DSVASLQ-K 1660

QY 229 QCLAVFSRVKFTRVLLTVLIAFT---KKETSAAVEAQKLMVQAADLLSAIHNSLHHGIQA 285

Db 1661 ECKIKSNTVESLQDVLTSVQARNAELEDEVSRSDKIRRRDDRCEHLSGKLLKLSQLEE 1720

QY 286 QNDT---TKGDHPIMMGF--EPLVNQ-RLLPPTFPYAKIKKEEMVNYFARLIDRIKTIV 339

Db 1721 QHETFFRAEQQMTQLGFLKETVKKQEKLLKLNLRQEQLI PRSSILVYESYIRIEKEI 1780

QY 340 CEVVNLNLHCILDFCFSESPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSF 399

Db 1781 --IVLQERLNGI-----ELSQQLP-----KGYFG----- 1802

QY 400 VDPVLPSPKCYLYNNHQA-KDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFA 458

Db 1803 -----YFFKTRNVEVLDSEFKQVAK-----LQFLAG---AEFIVKFKEDLEKCA 1845

QY 459 TLQDEFMTFYFNRAEKVD-----AALHTMLLKQEPQOQHLACLGTWVLYHNLR 507

Db 1846 AEEKEKQATFDNYSEKVENLGKSIEALYFALNREISPRKSLALSKS--AYHNLLV 1898

RESULT 15

TEL1\_YEAST

ID TEL1\_YEAST STANDARD; PRT; 2787 AA.

AC P38110;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Telomer length regulation protein TEL1.

GN TEL1 OR YBL088C OR YBL0706.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=96076635; PubMed=7502586;

RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;

RT "Sequence analysis of a 78.6 kb segment of the left end of

RT Saccharomycetes cerevisiae chromosome II.";

RL Yeast 11:1103-1112(1995).

RN [2]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RA Greenwell P.S., Kronmal S.L., Porter S.E., Gassenhuber J.,

RA Obermaier B., Petes T.D.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Putative phosphatidylinositol kinase involved in

CC controlling telomere length.

CC -!- SIMILARITY: Belongs to the PI3/PI4-kinase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X79489; CAA56016.1; -.

DR EMBL; Z35849; CAA84909.1; -.

DR EMBL; U31331; AAA69802.1; -.

DR PIR; S45416; S45416.

DR Germline; 138523; -.

DR SGD; S000184; TEL1.

DR GO; GO:0005634; C:nucleus; IC.

DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; IMP.

DR InterPro; IPR003151; FAT.

DR InterPro; IPR003152; FATC.

DR InterPro; IPR000403; PI3\_PI4\_kinase.

DR Pfam; PF02259; FAT; 1.

DR Pfam; PF02260; FATC; 1.

DR Pfam; PF00454; PI3\_PI4\_kinase; 1.

DR SMART; SM00146; PI3KC; 1.

DR PROSITE; PS00915; PI3\_4\_KINASE\_1; 1.

DR PROSITE; PS00916; PI3\_4\_KINASE\_2; 1.

DR PROSITE; PS0290; PI3\_4\_KINASE\_3; 1.

KW Transferase; Kinase.

FT DOMAIN 2461 2787 PI3K/PI4K.

FT CONFLICT 1190 1190 Y -> F (IN REF. 2).

SQ SEQUENCE 2787 AA; 321663 MW; 439B6E189E39499B CRC64;

Query Match 2.9%; Score 113; DB 1; Length 2787;

Best Local Similarity 22.4%; Pred. No. 9;

Matches 67; Conservative 50; Mismatches 96; Indels 86; Gaps 17;

QY 323 EEMVNYFARLIDRI-KTVCEV-VNLTNLHCILDFCFSESPCVLSRSLQTTFLVDNK 380

Db 1346 DEVITIFSSLLNLADKNTFEIEPSLPNLFCKIFIVLRENKQ----LSPSFQQAIKLLEHR 1401

QY 381 -----KVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQ---AKDC-IDSFV--- 425

Db 1402 DLIKIKTWKYFLDAIFGNIVQDD-----IYENTELLDDASDCGVDDVVLVS 1446

QY 426 ---THCVRPF-----CSLIQIHGHNRRARQDKLGHILEEFATLQDEFMTFYFNRAEKVDA 477

Db 1447 LLFSYARRPVASKIGCSLSKAAAINILKH-----HVPKEY--LSKNFKLWF-----A 1491

QY 478 ALHTMLLKQEPQOQHLACLGTWVLYHNLR-----MIQYLLSGF--EELYSMEHY 526

Db 1492 ALSRRILQOEVRERSTNFNNEVHLKNFEMVFRHPEQPHMIYQRISTFNKEAELYDSTEV 1551

QY 527 YVIYWLSEFLYAWLMS-TLSRADGSQMAEERIMEEQKGRSSKTKTKKKKKVRPLSREI 584

Db 1552 FFI---SECILTYLVGYSIGNSESEFCFRDNMNE-----NKDKVAPLDKDV 1595

Search completed: August 10, 2004, 21:50:35

Job time : 17 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:48:07 ; Search time 43 Seconds  
(without alignments)  
5378.485 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFDFSAHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archearp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3684	95.7	725	11 Q8K3H2	Q8k3h2 mus musculu
2	3648	94.8	725	11 Q9JI01	Q9ji01 rattus norv
3	3331.5	86.6	724	13 Q7T322	Q7t322 brachydanio
4	2670	69.4	526	4 Q9H703	Q9h703 homo sapien
5	1326	34.5	784	5 Q9W1A2	Q9w1a2 drosophila
6	1116	29.0	211	4 Q9H631	Q9h631 homo sapien
7	1017.5	26.4	708	5 Q8T3X5	Q8t3x5 drosophila
8	772	20.1	150	11 Q8BYJ9	Q8byj9 mus musculu
9	584	15.2	695	10 Q9SKH7	Q9skh7 arabidopsis
10	560.5	14.6	799	5 Q17003	Q17003 caenorhabdi
11	377.5	9.8	908	3 Q8WZW5	Q8wzw5 neurospora
12	252.5	6.6	708	3 Q9USY3	Q9usy3 schizosacch
13	190	4.9	41	11 Q8K3H1	Q8k3h1 mus musculu
14	150	3.9	1218	5 Q8I338	Q8i338 plasmodium
15	139.5	3.6	964	13 Q8JGR7	Q8jgr7 brachydanio
16	131	3.4	2965	5 Q8WRS5	Q8wrs5 plasmodium

17	130.5	3.4	893	5 Q8I3I6	Q8i3i6 plasmodium
18	129	3.4	2976	5 Q8WRS4	Q8wrs4 plasmodium
19	127	3.3	733	11 Q8VI24	Q8vi24 mus musculu
20	127	3.3	969	16 Q9ZJA3	Q9zja3 helicobacte
21	126	3.3	2770	5 Q7YUE9	Q7yue9 plasmodium
22	126	3.3	2957	5 Q8WRS6	Q8wrs6 plasmodium
23	123.5	3.2	778	5 Q19642	Q19642 caenorhabdi
24	122.5	3.2	728	11 Q8K454	Q8k454 mus musculu
25	122.5	3.2	1607	5 Q20187	Q20187 caenorhabdi
26	122.5	3.2	2253	13 P70012	P70012 xenopus lae
27	122	3.2	1175	16 Q8XNW6	Q8xnw6 clostridium
28	121	3.1	966	11 Q8C3A9	Q8c3a9 mus musculu
29	121	3.1	966	11 Q7TQK1	Q7tqk1 mus musculu
30	119	3.1	986	4 Q8TBW2	Q8tbw2 homo sapien
31	119	3.1	1196	4 Q94834	Q94834 homo sapien
32	119	3.1	3412	4 Q60316	Q60316 homo sapien
33	118.5	3.1	550	11 Q8BIZ2	Q8biz2 mus musculu
34	118.5	3.1	1029	10 Q8O955	Q8o955 arabidopsis
35	118.5	3.1	5303	5 Q9V628	Q9v628 drosophila
36	118	3.1	728	4 Q60561	Q60561 homo sapien
37	118	3.1	793	5 Q8I5C9	Q8i5c9 plasmodium
38	116.5	3.0	1045	10 Q94G20	Q94g20 daucus caro
39	116.5	3.0	1641	11 Q88528	Q88528 mus musculu
40	115.5	3.0	550	11 Q8BIL5	Q8bil5 mus musculu
41	115	3.0	721	5 Q27308	Q27308 anopheles a
42	115	3.0	1275	4 Q8TE30	Q8te30 homo sapien
43	115	3.0	1316	10 Q7Y239	Q7y239 oryzaativ
44	114	3.0	546	12 Q98682	Q98682 simian cyto
45	114	3.0	865	6 Q95LL1	Q95ll1 macaca fasc

ALIGNMENTS

RESULT 1

Q8K3H2 ID Q8K3H2 PRELIMINARY; PRT; 725 AA.  
AC Q8K3H2;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Embryonic growth-associated protein EGAP.  
GN C030004C14RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wenzlau J.M., Weiser-Evans M.C.M.;  
RT "Mouse Embryonic Growth Associated Protein (EGAP).";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBBJ databases.  
DR EMBL; AY102701; AAM52342.1; --  
DR MGD; MGI:1925939; C030004C14Rik.  
DR InterPro; IPR007244; Mak10.  
DR Pfam; PF04112; Mak10; 1.  
SQ SEQUENCE 725 AA; 83349 MW; F9C7E798F7B51DFC CRC64;

Query Match 95.7%; Score 3684; DB 11; Length 725;  
Best Local Similarity 95.9%; Pred. No. 2.6e-300;  
Matches 703; Conservative 11; Mismatches 11; Indels 8; Gaps 1;

QY	1	MVMKASVDDDDSGWELSMPEKMSNTNWVDITQDFEEACRELKLGELLDKLFGLFEAM	60
		:     :     :     :     :     :     :     :	
Db	1	MVMKATVDDDDASGWELGVPEKMSSTSWVDITQDFEDACRELKLGELLDKLFGLFEAM	60
QY	61	SAIEMMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLTLPELIGIMDTFCCLITWL	120
		:     :     :     :     :     :     :     :	
Db	61	SAIEMMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLSLPELIGIMDTFCCLITWL	120
QY	121	EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY	180
		:     :     :     :     :     :     :     :	
Db	121	EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY	180



QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDPEVELEHQOCLAVFSRVKFT 240  
Db 181 GFKMANSVTDLRVTGMLKDVEDDLQRRVKSTRSQGEERDPEVELEHQOCLAVFSRVKFT 240  
QY 241 RVLLTVLIAFTKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPIMMGF 300  
Db 241 RVLLTVLIAFTKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNGTTKGDPIMMGF 300  
QY 301 EPLVNQRLLPPTFPKYAKIKREEMVNYFARLIDRIKTVCEVWNLNLHLCILDFCFE 360  
Db 301 EPLVNQRLLPPTFPKYAKIKREEMVNYFARLIDRIKTVCEVWNLNLHLCILDFCFE 360  
QY 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
Db 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFTFYFNRAEKVDAALH 480  
Db 421 IDSFVTHCVRPFCSLVQIHGHNRRARQDKLGHILEEFATLQDEFTFYFNRAEKVDAALH 472  
QY 481 TMLLKQEPQORHLACLTGTVWLYHNLRIIMIQYLLSGFELELYSMHEYYIYWYLSEFLYAW 540  
Db 473 TMLLKQEPQORHLACLTGTVWLYHNLRIIMIQYLLSGFELELYSMHEYYIYWYLSEFLYAW 532  
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNNMCAGMFKT 600  
Db 533 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNNMCAGMFKT 592  
QY 601 MVAFDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 660  
Db 593 MVAFDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 652  
QY 661 PELYVAASKHFQQAAMILENIPNDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Db 653 PELYVAASKHFQQAAMILENIPNDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 712  
QY 721 FSAHKYFPVVKLV 733  
Db 713 FSVHKYFPVVKLV 725  
RESULT 2  
Q9JI01 PRELIMINARY; PRT; 725 AA.  
AC Q9JI01  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Corneal wound healing related protein.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20314365; PubMed=10855038;  
RA Yi X.J., Li X.F., Yu F.S.;  
RT "A novel epithelial wound-related gene is abundantly expressed in  
RT developing rat cornea and skin."  
RL Curr. Eye Res. 20:430-440(2000).  
DR EMBL; AF272892; AAF81791.1; -.  
DR InterPro; IPR007244; Mak10.  
DR Pfam; PF04112; Mak10; 1.  
SQ SEQUENCE 725 AA; 83151 MW; ADB38695BDE1CD2D CRC64;

Query Match 94.8%; Score 3648; DB 11; Length 725;  
Best Local Similarity 95.0%; Pred. No. 2.7e-297;  
Matches 696; Conservative 15; Mismatches 14; Indels 8; Gaps 1;  
QY 1 MVMKASVDDDDSGWELSMPEKMEKSNWVDITQDFEEACBELKIGELLHDKLFLGLFEAM 60  
Db 1 MVMKAAVDDDDASGWELNVPEKMEKSTSWVDITQDFEDACBELKIGELLHDKLFLGLFEAM 60

QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDITLPELIGIMDTCFCCLITWL 120  
Db 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIVKDKGTIKIKDLSLPELIGIMDTCFCCLITWL 120  
QY 121 EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMTY 180  
Db 121 EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMTY 180  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDPEVELEHQOCLAVFSRVKFT 240  
Db 181 GFKMANGVTDLRVTGMLKDVEDDMQRRVKSTRSQGEERDPEVELEHQOCLAVFSRVKFT 240  
QY 241 RVLLTVLIAFTKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPIMMGF 300  
Db 241 RVLLTVLIAFTKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNGTTKGDPIMMGF 300  
QY 301 EPLVNQRLLPPTFPKYAKIKREEMVNYFARLIDRIKTVCEVWNLNLHLCILDFCFE 360  
Db 301 EPLVNQRLLPPTFPKYAKIKREEMVNYFARLIDRIKTVCEVWNLNLHLCILDFCFE 360  
QY 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
Db 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQDEFTFYFNRAEKVDAALH 480  
Db 421 IDSFVTHCVRPFCSLVQIHGHNRRARQDKLGHILEEFATLQDEFTFYFNRAEKVDAALH 472  
QY 481 TMLLKQEPQORHLACLTGTVWLYHNLRIIMIQYLLSGFELELYSMHEYYIYWYLSEFLYAW 540  
Db 473 TMLLKQEPQORHLACLTGTVWLYHNLRIIMIQYLLSGFELELYSMHEYYIYWYLSEFLYAW 532  
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNNMCAGMFKT 600  
Db 533 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNNMCAGMFKT 592  
QY 601 MVAFDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 660  
Db 593 MVAFDMDGKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 652  
QY 661 PELYVAASKHFQQAAMILENIPNDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Db 653 PELYVAASKHFQQAAMILENIPNDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 712  
QY 721 FSAHKYFPVVKLV 733  
Db 713 FSVHKYFPVVKLV 725

RESULT 3  
Q7T322 PRELIMINARY; PRT; 724 AA.  
AC Q7T322  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,





```
RESULT 5
Q9W1A2      PRELIMINARY;      PRT;      784 AA.
AC Q9W1A2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG4065 protein (GM08606P).
GN CG4065.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003463; AAF47171.1; -.
DR EMBL; AY060948; AAL28496.1; -.
DR FlyBase; FBgn0034982; CG4065.
DR InterPro; IPR007244; Mak10.
DR Pfam; PF04112; Mak10; 1.
SQ SEQUENCE 784 AA; 89204 MW; DBDB966F0F523F17 CRC64;
Query Match 34.5%; Score 1326; DB 5; Length 784;
```

```
Best Local Similarity 39.3%; Pred. No. 2.7e-102;
Matches 291; Conservative 131; Mismatches 246; Indels 72; Gaps 14;

QY 29 WVDITQDFEEACRELKGLGELLHDKLFGLFEMSALFEMMDPKMDAGMIGNQVNRKVLNFEQ 88
Db 80 WVDVTKEFHDACAELOPGEAQDMLFGLFEAMSAIEIMDPKMDVGMGDFKQDLPPSFEA 139
QY 89 AIKDGTIKIKDLTLPDLIGIMDTCFCCCLITWLEGHSLAQTVFTCLYIHNPDFIEDPAMKA 148
Db 140 AIATGAIKLDDLTPSELIGIYDALFSLVSWLEGNMDQVLTCLYLHAPAQIKDKALRV 199
QY 149 FALGILKICDIAREKVNKAAFVEEEDFQSMYTG---FKMANSVTDLRVTGMLKDVEDDMQ 205
Db 200 FCTAVRNLIIVIKIIAIAAANVEEEDFQ--LYGNSALLAAEKAQATVYSSLDKDVDELI 257
QY 206 RRVKSTRSRQGEERDEVELEHQQLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAQ--- 262
Db 258 RKCKKLTST-----EDWMAVVHRLRFMRHLFQVIYHVEQMASNDTVDDKVDI 304
QY 263 -KLMVQAADLLSAIHNSLHHGIAQNDDTTKG-DHPIMMGFEPLVNQRLLPPTFPFYAKII 320
Db 305 YKILLVASEMLPGIRNLTDRGTQPE---KGSDAENPMGFSPRIHDRSQPPAFPRSIKIR 360
QY 321 KREEMVNYFARLIDRIKTVCEVWNLTNLHCILDFCFEFSEQS-PCVLSRSLLTQTTFLVDN 379
Db 361 DRPSSYQFLEEMISRFKYACKVTYKYDYYSALNFFIEYSKSGQCILSRSLVQLTFLSANM 420
QY 380 KKVFGLHMQDMVKDALRSFVDPVLPSPKCYLYNNEHQAKDCIDSFVTHCV--RPFCSLIQ 437
Db 421 RMAHGKLPKQFLRHSVQVFNPPVLNAKHPVAADPKVQOHLNFFRYCINMNTFTQFIR 480
QY 438 IHGHNRRQRDKLGHILEEFPATLQDEFMTFYFNRAEKYDAALHTMLLKQ-----E 487
Db 481 ICGFNRRQRDKLARLIENFDTIQ-----VDAARLDSMMNQLANERAMEGNE 527
QY 488 PQR---QHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYYIYWLSEFLYAWLMST 544
Db 528 PMATALKHSTHFTWVLYNCFRAMLIFLMSGFELELYAVHEFLYIYWYPYEFLLGLVSA 587
QY 545 LSRADGSQMAEERIMEEQQGRS-----SKTKKKKKVRPLSREITMSQAYQNM 593
Db 588 LTRTENILLAQEEYAEHQSKTQSGSGGAACKNRKAAKPKNKKTQRPYRAEIVFYHALLSL 647
QY 594 CAGMFKTMVAFDMGDKVRKPKFELDSEQVRYEHRFAPENSVMTPPPVHYLQFKEMSD-IN 652
Db 648 CGGMVKAMGALTQDGRVLPFLSKFDNEEIRYNNRFLPATLTSPPPVSYAEFKNVREHM 707
QY 653 KYSPPPPQSPELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKKE 712
Db 708 RHS---VEDLYTYAAKHFDQARNVLESIQNPQDQMLDLQLIARTNFVVMNVLARGHKE 763
QY 713 SKVPPPEDFSAHKYFPVVKL 732
Db 764 VKRQPEDFDSKHSYFPIIKL 783
```

RESULT 6

```
Q9H631      PRELIMINARY;      PRT;      211 AA.
ID Q9H631
AC Q9H631;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ22643.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
```















Qy 682 ---PNPDHEVNRI-LKVA-----KP--NFVVMKLLAGGHKKESKVPPEFDFSAHKY 726  
Db 838 GSSPGLFRKIQAVCLKVSSTLQTKPGSDFKI-PLESKTNEIEQKVEPHNDYFSTQF 892

Search completed: August 10, 2004, 21:51:28  
Job time : 47 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:44:16 ; Search time 58 Seconds  
(without alignments)  
3570.816 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEDFSAHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3849	100.0	733	5	ABP52902 Human lun
2	3783	98.3	725	5	AAO19400 Human mol
3	3728.5	96.9	726	7	ADC31012 Human nov
4	1326	34.5	784	4	ABB60388 Drosophil
5	1100	28.6	239	5	ABG69816 Human sec
6	685	17.8	160	7	ADC32736 Human nov
7	127	3.3	733	3	AAB07449 Amino aci
8	127	3.3	733	3	AAB07448 Amino aci
9	127	3.3	761	6	AAE36464 Human col
10	127	3.3	761	7	ADD14191 Human src
11	127	3.3	1089	2	AAW01896 Nonsense-
12	127	3.3	1089	2	AAY05835 Yeast Nmd
13	127	3.3	1089	3	AAY98055 Yeast Nmd
14	126	3.3	1257	4	ABG17335 Novel hum
15	121.5	3.2	1184	4	ABG02338 Novel hum
16	120	3.1	644	4	AAB95305 Human pro
17	119	3.1	1488	7	ADC31044 Human nov
18	119	3.1	3353	5	AAU82706 Amino aci
19	119	3.1	3412	7	ADB68580 Human ubi
20	118.5	3.1	5303	4	ABB67866 Drosophil
21	116	3.0	736	5	ABP65160 Hypoxia-r
22	115.5	3.0	1141	4	ABG17210 Novel hum
23	115	3.0	633	5	ABG96286 Human ova
24	115	3.0	912	4	AAB90771 Human she
25	115	3.0	912	5	ABG34124 Human non

26	115	3.0	912	6	ABU89754	Abu89754 Protein d
27	115	3.0	937	4	AAB63240	Aab63240 Human bre
28	115	3.0	1187	5	ABP52168	Abp52168 Mouse pot
29	115	3.0	1187	7	ADD37507	Add37507 Mouse pho
30	115	3.0	1187	7	ADD37469	Add37469 Mouse pho
31	115	3.0	1565	4	ABG06288	Abg06288 Novel hum
32	114.5	3.0	1029	4	ABG15642	Abg15642 Novel hum
33	114.5	3.0	1190	4	ABG23670	Abg23670 Novel hum
34	114.5	3.0	1190	4	ABG06619	Abg06619 Novel hum
35	114.5	3.0	1190	4	ABG05642	Abg05642 Novel hum
36	114.5	3.0	1191	6	ABJ19343	Abj19343 NOVX rela
37	114.5	3.0	1304	4	ABG14789	Abg14789 Novel hum
38	114	3.0	985	4	ABG06938	Abg06938 Novel hum
39	114	3.0	1065	4	ABG02332	Abg02332 Novel hum
40	114	3.0	1078	4	ABG10803	Abg10803 Novel hum
41	114	3.0	1078	4	ABG06060	Abg06060 Novel hum
42	114	3.0	1078	4	ABG06606	Abg06606 Novel hum
43	114	3.0	1078	4	ABG09642	Abg09642 Novel hum
44	114	3.0	1261	4	ABG07064	Abg07064 Novel hum
45	114	3.0	1558	3	AAB18324	Aab18324 Plasmodiu

ALIGNMENTS

RESULT 1  
ABP52902  
ID ABP52902 standard; protein; 733 AA.  
XX  
AC ABP52902;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Human lung specific protein sequence SEQ ID NO:145.  
XX  
KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein;  
KW LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;  
KW squamous cell carcinoma.  
XX  
OS Homo sapiens.  
XX  
PN WO200264788-A2.  
XX  
PD 22-AUG-2002.  
XX  
PF 20-NOV-2001; 2001WO-US045080.  
XX  
PR 20-NOV-2000; 2000US-0252054P.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Macina RA, Recipon H, Chen S, Sun Y, Liu C;  
XX  
DR WPI; 2002-657601/70.  
XX  
PT New lung specific nucleic acid useful in gene therapy or as vaccines for  
PT treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung  
PT diseases, as well as for diagnosing, monitoring or staging these  
PT diseases.  
PT  
XX Claim 11; Page 239-242; 282pp; English.  
XX  
PS  
XX  
CC The present invention describes an isolated lung specific nucleic acid  
CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid  
CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;  
CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),  
CC given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);  
CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific  
CC protein (LSP) sequences have cytostatic activity and can be used in gene  
CC therapy and vaccines. LSNA and LSPs are useful for diagnosing and  
CC monitoring the presence and metastases of lung cancer in a patient. An  
CC antibody that specifically binds to an LSP can be used for determining  
CC the presence of an LSP in a sample, as well as for treating a patient

CC with lung cancer, particularly by inducing an immune response against the  
CC lung cancer cell expressing the LSNAs or LSPs. In particular, these LSNAs  
CC and LSPs are useful for identifying, diagnosing, monitoring, staging,  
CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-  
CC cancerous disease states in lung  
XX  
SQ Sequence 733 AA;

Query Match 100.0%; Score 3849; DB 5; Length 733;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLEFAM 60  
Db |||||  
1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLEFAM 60  
QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLTLPeligimdtcfccclitwl 120  
Db |||||  
61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLTLPeligimdtcfccclitwl 120  
QY 121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMTY 180  
Db |||||  
121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMTY 180  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQCLAVFSRVKFT 240  
Db |||||  
181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQCLAVFSRVKFT 240  
QY 241 RVLTLVLIATFKKETSAAVEAQKLMVQAADLLSAIHNSLHGIIQAQNDTTKGDHPIMMGF 300  
Db |||||  
241 RVLTLVLIATFKKETSAAVEAQKLMVQAADLLSAIHNSLHGIIQAQNDTTKGDHPIMMGF 300  
QY 301 EPLVNQRLLPPTFPYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCILDFCFESE 360  
Db |||||  
301 EPLVNQRLLPPTFPYAKIIKREEMVNYFARLIDRIKTVCEVNLNLHCILDFCFESE 360  
QY 361 QSPCVLSRLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPSPKCYLYNNHQAKDC 420  
Db |||||  
361 QSPCVLSRLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVPSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFFNRAEKVDAALH 480  
Db |||||  
421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFFNRAEKVDAALH 480  
QY 481 TMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540  
Db |||||  
481 TMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540  
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCAGMFKT 600  
Db |||||  
541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCAGMFKT 600  
QY 601 MVAFMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 660  
Db |||||  
601 MVAFMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPPPQS 660  
QY 661 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Db |||||  
661 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
QY 721 FSAHKYFPVVKLV 733  
Db |||||  
721 FSAHKYFPVVKLV 733

RESULT 2  
AAO19400  
ID AAO19400 standard; protein; 725 AA.  
XX  
AC AAO19400;  
XX  
DT 10-DEC-2002 (first entry)  
XX

DE  
XX  
KW  
KW  
KW  
XX  
OS  
XX  
PN  
XX  
PD  
XX  
PF  
XX  
PR  
PR  
PR  
PR  
PR  
PR  
PR  
PR  
XX  
PA  
XX  
PI  
PI  
PI  
PI  
PI  
PI  
XX  
DR  
DR  
XX  
PT  
PT  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

Query Match  
Best Local Similarity  
Matches 724; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLEFAM 60  
Db |||||  
1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLEFAM 60  
QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLTLPeligimdtcfccclitwl 120  
Db |||||  
61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDLTLPeligimdtcfccclitwl 120  
QY 121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMTY 180  
Db |||||  
121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMTY 180  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQCLAVFSRVKFT 240  
Db |||||  
181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQCLAVFSRVKFT 240

Human molecule for disease detection and treatment protein #3.  
Human; molecule for disease detection and treatment; MDDT; gene therapy;  
cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;  
antiinflammatory; antiasthmatic; cerebroprotective; nootropic;  
neuroprotective; antiparkinsonian; cardiant; antiangular.

Homo sapiens.  
WO200270709-A2.  
12-SEP-2002.  
08-FEB-2002; 2002WO-US003709.  
09-FEB-2001; 2001US-0268117P.  
15-FEB-2001; 2001US-0269618P.  
23-FEB-2001; 2001US-0271118P.  
07-MAR-2001; 2001US-0274486P.  
09-MAR-2001; 2001US-0274436P.  
28-NOV-2001; 2001US-0334229P.  
01-FEB-2002; 2002US-0353284P.

(INCY-) INCYTE GENOMICS INC.

Lal PG, Baughn MR, Yao MG, Walia NK, Elliot VS, Xu Y;  
Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;  
Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;  
Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;  
Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;

WPI; 2002-713453/77.  
N-PSDB; AAL49929.

New human molecules for disease detection and treatment (MDDT), useful  
for diagnosing, treating and preventing diseases or conditions associated  
with the aberrant MDDT expression, e.g. cancer, AIDS, asthma, diabetes,  
hepatitis.

Claim 1; Page 128-130; 177pp; English.

The present invention relates to human proteins and coding sequences of  
molecules for disease detection and treatment MDDT. The sequences can be  
used in the treatment of diseases associated with the decreased  
expression or overexpression of MDDT, such as cell proliferative (cancer,  
atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,  
allergies, Addison's disease, asthma), developmental (dwarfism, renal  
tubular acidosis), neurological (e.g. stroke, Parkinson's disease,  
epilepsy) and cardiovascular (congestive heart failure, myocardial  
infarction, angina pectoris) disorders. The present sequence is a protein  
of the invention

Sequence 725 AA;

QY 241 RVLTLVLIATFKKETS AVAEAKLMVQAADLLSAIHNSLHHG IQAQNDDTTKG DHPIMMGF 300  
Dbb 241 RVLTLVLIATFKKETS AVAEAKLMVQAADLLSAIHNSLHHG IQAQNDDTTKG DHPIMMGF 300  
QY 301 EPLVNQRLLPPTFPFYAKIIKKEEMVNYFARLIDRIKTVCEVWNLTNLHCILDFCFEFS 360  
Dbb 301 EPLVNQRLLPPTFPFYAKIIKKEEMVNYFARLIDRIKTVCEVWNLTNLHCILDFCFEFS 360  
QY 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAKDC 420  
Dbb 361 QSPCVLSRSLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIQIHGHNARQORDKLGHILEEFATLQDEFFMTFFYNRAEKVDAALH 480  
Dbb 421 IDSFVTHCVRPFCSLIQIHGHNARQORDKLGHILEEFATLQDE-----AEKVDAALH 472  
QY 481 TMLLKQEPQORQHLACLTWVLYHNLRLIMIQYLLSGFELELYSMHEYYYIYWYLSEFLYAW 540  
Dbb 473 TMLLKQEPQORQHLACLTWVLYHNLRLIMIQYLLSGFELELYSMHEYYYIYWYLSEFLYAW 532  
QY 541 LMSLTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSRITMSQAYQNNMCAGMEKT 600  
Dbb 533 LMSLTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSRITMSQAYQNNMCAGMEKT 592  
QY 601 MVAFDMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQPKEMSDLNKYSPPPOS 660  
Dbb 593 MVAFDMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQPKEMSDLNKYSPPPOS 652  
QY 661 PELYYAASKHFQQA KMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Dbb 653 PELYYAASKHFQQA KMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 712  
QY 721 FSAHKYFPVVKLV 733  
Dbb 713 FSAHKYFPVVKLV 725  
RESULT 3  
ADC31012  
ID ADC31012 standard; protein; 726 AA.  
XX  
AC ADC31012;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human novel polypeptide sequence, SEQ ID NO:1094.  
XX  
KW Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 9.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029271-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 24-SEP-2002; 2002WO-US030474.  
XX  
PR 24-SEP-2001; 2001US-0324631P.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;

DR WPI; 2003-371981/35.  
DR N-PSDB; ADC30041.  
XX  
PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
treating conditions such as neurodegenerative diseases, anemias, platelet  
disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
cancer.  
PT  
XX  
XX Claim 20; SEQ ID NO 1094; 1185pp; English.  
XX  
PS The invention relates to 971 novel human cDNA sequences (ADC29919-  
ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
invention also relates to nucleic acid sequences over 99% identical with  
the novel human cDNAs. The invention additionally encompasses expression  
vectors and host cells comprising a nucleic acid of the invention; the  
recombinant production of a polypeptide of the invention; an antibody  
against a polypeptide of the invention; a method of detecting  
polynucleotides or polypeptides of the invention; and methods of  
identifying a compound which binds to a polypeptide of the invention. The  
invention further discloses methods of preventing, treating or  
ameliorating a medical condition; kits comprising polynucleotide probes  
and/or monoclonal antibodies for carrying out the methods of the  
invention; methods for the identification of compounds that modulate the  
expression or activity of the polynucleotide and/or polypeptide; and 767  
contig sequences corresponding to the cDNA sequences of the invention  
(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
-ADC33394). The nucleic acids and polypeptides of the invention are  
useful in diagnostics, drug screening, forensics, gene mapping, in the  
identification of mutations responsible for genetic disorders or other  
traits, for assessing biodiversity, and in producing many other types of  
data and products dependent on DNA and amino acid sequences. They are  
also used for treating diseases such as Parkinson's disease, Alzheimer's  
disease and other neurodegenerative diseases, anaemia, platelet  
disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
cancer. The nucleic acids may also be used as hybridisation probes or  
primers, and in the recombinant production of a protein. The polypeptides  
are also useful in generating antibodies, as molecular weight markers,  
and as food supplements. The present sequence represents a specifically  
claimed human polypeptide sequence of the invention. Note: The sequence  
data for this patent did not form part of the printed specification, but  
was obtained in electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 726 AA;  
Query Match 96.9%; Score 3728.5; DB 7; Length 726;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 718; Conservative 1; Mismatches 6; Indels 9; Gaps 2;  
QY 1 MVMKASVDDDDSGWELSMPEKMSNTNWVDITQDFEEACRELKLGELLHDKLFLGLFEAM 60  
Dbb 1 MVMKASVDDDDSGWELSMPEKMSNTNWVDITQDFEEACRELKLGELLHDKLFLGLFEAM 60  
QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQA IKDGTIKIKDLTLP ELIGIMDTCFCLITWL 120  
Dbb 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQA IKDGTIKIKDLTLP ELIGIMDTCFCLITWL 120  
QY 121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAPALGILKICDIAREKVNKAAVFEEDFQSMTY 180  
Dbb 121 EGHSLAQTVFTCLYIHNPDFIEDPAMKAPALGILKICDIAREKVNKAAVFEEDFQSMTY 180  
QY 181 GFKMANSVTDLRVGTGMLKDVEDDMQRRVVKSTRSRQGEERDPEVELEHQQLAVFSRVKFT 240  
Dbb 181 GFKMANSVTDLRVGTGMLKDVEDDMQRRVVKSTRSRQGEERDPEVELEHQQLAVFSRVKFT 240  
QY 241 RVLLTVLIAFTKKETS AVAEAKLMVQAADLLSAIHNSLHHG IQAQNDDTTKG DHPIMMGF 300  
Dbb 241 RVLLTVLIAFTKKETS AVAEAKLMVQAADLLSAIHNSLHHG IQAQNDDTTKG DHPIMMGF 300  
QY 301 EPLVNQRLLPPTFPFYAKIIKKEEMVNYFARLIDRIKTVCEVWNLTNLHCILDFCFEFS 360  
Dbb 301 EPLVNQRLLPPTFPFYAKIIKKEEMVNYFARLIDRIKTVCEVWNLTNLHCILDFCFEFS 360



QY 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVD-PPVLSPKCYLYNNHQAKD 419  
Dbb 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRVFCQILRVLSPKCYLYNNHQAKD 420  
QY 420 CIDSFVTHCVRPFCSLIQIHGHNRRQRDKLGHILEEFATLQDEFFMTFYFNRAEKVDAAL 479  
Dbb 421 CIDSFVTHCVRPFCSLIQIHGHNRRQRDKLGHILEEFATLQDE-----AEKVDAAL 472  
QY 480 HTMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGFELEYSMHEYYYIYWYLSEFLYA 539  
Dbb 473 HTMLLKQEPQRQHLAWLGTWVLYHNLRIMIQYLLSGFELEYSMHEYYYIYWYLSEFLYA 532  
QY 540 WLMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGMEK 599  
Dbb 533 WLMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGMEK 592  
QY 600 TMVAFDMGDKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLOFKEMSDLNKYSPPPQ 659  
Dbb 593 TMVAFDMGDKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLOFKEMSDLNKYSPPPQ 652  
QY 660 SPELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEF 719  
Dbb 653 SPELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEF 712  
QY 720 DFSAHKYFPVVKLV 733  
Dbb 713 DFSAHKYFPVVKLV 726

RESULT 4  
ABB60388 standard; protein; 784 AA.  
ID ABB60388  
XX AC ABB60388;  
XX 26-MAR-2002 (first entry)  
Drosophila melanogaster polypeptide SEQ ID NO 7956.

Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US009231.

23-MAR-2000; 2000US-0191637P.

11-JUL-2000; 2000US-00614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL04491.

New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signaling and cell-cell  
interactions.

Disclosure; SEQ ID NO 7956; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 784 AA;  
Query Match 34.5%; Score 1326; DB 4; Length 784;  
Best Local Similarity 39.3%; Pred. No. 2.4e-122;  
Matches 291; Conservative 131; Mismatches 246; Indels 72; Gaps 14;  
QY 29 WVDITQDFEEACRELKLGELLHDKLFGLEFAMSAMMDPKMDAGMIGNQVNRKVLNFEQ 88  
Dbb 80 WVDVTKFEHDACAELQPGELAQDMLFGLFEAMSAMMDPKMDVGMGFDKQDLPPPSFEA 139  
QY 89 AIKDGTIKIKDLTLPELIGIMDTFCCLITWLEGHSLAQTVFTCLYIHNPDFIEDPAMKA 148  
Dbb 140 AIATGAIKLDDLTPSELIGIYDALFSLVSWLEGNSMDQVLTCLYLHAPAIKDKALRV 199  
QY 149 FALGILKICDIAREKVNKAAPVEEEDFQSMYTG---FKMANSVTDLRVTGMLKQVEDDMQ 205  
Dbb 200 FCTAVRNLIIVIKKIIAIAVAVNEEDFQ--LYGNSALLAAEKAQATVYSSLKQVDELI 257  
QY 206 RRVKSTRSRQGEERDPEVELEHQQLAVRSRVKFTRVLLTVLIAFTKKTSAVAEAQ--- 262  
Dbb 258 RKCKKLST-----EDWMAVVRHRLRFMRHLFQVIYHVEQMASNDTVDDKVDI 304  
QY 263 -KLMVQAADLLSAIHNSLHHGIAQNDTTKG-DHPIMMGFEPLVNQRLLPPTFPYAKII 320  
Dbb 305 YKILLVASEMLPGIRNTLDRGTQPE---KGDAPNPMGFSPRIHDRSQPPAFPRSIKIR 360  
QY 321 KREEMVNYFARLIDRIKTVCEVNVNLTNLHCILDFCFESEQS-PCVLSRSLLOTTFLVDN 379  
Dbb 361 DRPSSYQFLEEMISRFKACKVTYKYDYSAALNFFIEYSKSGQCILSRSVLQTLFSANM 420  
QY 380 KKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAKDCIDSFVTHCV--RPFCSLIQ 437  
Dbb 421 RMAHGKLPKQFLRHSVQVFNPPVLNAKHPVAADPKVQOQHLENNFRYCINMTTQFIR 480  
QY 438 IHGHNRRQRDKLGHILEEFATLQDEFFMTFYFNRAEKVDAALHTMLLKQ-----E 487  
Dbb 481 ICGFNRRQRDKLARLIENFDTIQ-----VDAARLDSMMNQLANERAMEGNE 527  
QY 488 PQR---QHLACLGTWVLYHNLRIMIQYLLSGFELEYSMHEYYYIYWYLSEFLYAWLMST 544  
Dbb 528 PMATALKSHSTHFSTWVLYNCFRAMLIFLMSGFELEYAVHEFLYIYWYPYEFLLGLVSA 587  
QY 545 LSRADGSQMAEERIMEEQKGRS-----SKTKKKKKVRPLSREITMSQAYQNM 593  
Dbb 588 LTRTENILLAQEEYAEHQSKTQSGGGAANKRKAAPKKNKKTQRPYRAEIVFYHALLSL 647  
QY 594 CAGMFKTMVAFDMGDKVRKPKFELDSQVRYEHRFAPFNSVMTPPPVHYLOFKEMSD-LN 652  
Dbb 648 CGMYKAMGALT KDGRVRLPLSKFDNEEIRYNRRFLPFATLTSPPPVSYAEFFKNVREHM 707  
QY 653 KYSPPQPSPELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKKE 712  
Dbb 708 RHS---VEDLYTYAAKHDFDQARNVLESIQNPDOEMDLDLQIARTNFVVMNVLARGHKE 763  
QY 713 SKVPPPEFDSAHHKYFPVVKLV 732  
Dbb 764 VKRQPEFDSKHSYFPIKL 783

RESULT 5  
ABG69816 standard; protein; 239 AA.  
ID ABG69816  
XX AC ABG69816;  
XX 21-OCT-2002 (first entry)  
XX Human secretory protein #7.  
DE  
XX

KW Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;  
KW bursitis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;  
KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;  
KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;  
KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;  
KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;  
KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;  
KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;  
KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;  
KW Parkinson's disease; central nervous system disorder; mental disorder;  
KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;  
KW gene therapy.  
XX  
OS Homo sapiens.  
XX WO200257304-A2.  
PN  
XX  
PD 25-JUL-2002.  
XX  
PF 15-JAN-2002; 2002WO-US001340.  
XX  
PR 16-JAN-2001; 2001US-0261864P.  
PR 16-JAN-2001; 2001US-0261865P.  
PR 16-JAN-2001; 2001US-0261979P.  
PR 16-JAN-2001; 2001US-0261981P.  
PR 17-JAN-2001; 2001US-0262164P.  
PR 17-JAN-2001; 2001US-0262208P.  
PR 17-JAN-2001; 2001US-0263131P.  
PR 19-JAN-2001; 2001US-0262599P.  
PR 19-JAN-2001; 2001US-0262760P.  
PR 19-JAN-2001; 2001US-0263063P.  
PR 19-JAN-2001; 2001US-0263066P.  
PR 19-JAN-2001; 2001US-0263069P.  
PR 19-JAN-2001; 2001US-0263070P.  
PR 19-JAN-2001; 2001US-0263074P.  
PR 19-JAN-2001; 2001US-0263076P.  
PR 19-JAN-2001; 2001US-0263077P.  
PR 19-JAN-2001; 2001US-0263329P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;  
PI Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;  
PI Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;  
XX  
DR WPI; 2002-590716/63.  
DR N-PSDB; ABS51309.  
XX  
PT New purified secretory polypeptides and polynucleotides, useful in the  
PT diagnosis, study, prevention or treatment of diseases associated with  
PT decreased expression of functional secretory molecules, e.g. AIDS, cancer  
PT or allergies.  
XX  
PS Claim 27; Page 304-305; 340pp; English.  
XX  
CC The invention describes an isolated polynucleotide a naturally occurring  
CC polynucleotide sequence at least 90 % identical to it, a polynucleotide  
CC complementary to it or an RNA equivalent of it. The purified secretory  
CC polypeptides (SPTM) and polynucleotides are useful in the diagnosis,  
CC study, prevention or treatment of diseases associated with decreased  
CC expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,  
CC bursitis, cirrhosis, hepatitis, polycythaemia vera, primary  
CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,  
CC leukaemia, myeloma or sarcoma, immune system disorder such as acquired  
CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,  
CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,  
CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,  
CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid  
CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's  
CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's  
CC disease, other developmental disorder of the central nervous system,  
CC mental disorder including mood, anxiety or schizophrenic disorder,  
CC amnesia or Tourette's disorder. The polynucleotides may be used in

CC hybridisation and amplification technologies, e.g. in assessing gene  
CC expression patterns, to develop a transcript image for a particular cell  
CC or tissue, or to create transgenic animals to model human disease. This  
CC is the amino acid sequence of a human secretory protein isolated in the  
CC invention  
XX  
SQ Sequence 239 AA;  
  
Query Match 28.6%; Score 1100; DB 5; Length 239;  
Best Local Similarity 99.5%; Pred. No. 1.2e-100;  
Matches 211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACBELKLGLLHDKLFLGLEAM 60  
|||||  
Db 28 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACBELKLGLLHDKLFLGLEAM 87  
  
Qy 61 SAITEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTLPGLIGIMDTFCCLITWL 120  
|||||  
Db 88 SAITEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTLPGLIGIMDTFCCLITWL 147  
  
Qy 121 EGHSLAQTVFTCLYIHNPFIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMTY 180  
|||||  
Db 148 EGHSLAQTVFTCLYIHNPFIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMTY 207  
  
Qy 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTR 212  
|||||  
Db 208 GFKMANSVTDLRVTGMLKDVGDMDQRRVKSTR 239  
  
RESULT 6  
ADC32736  
ID ADC32736 standard; protein; 160 AA.  
XX  
AC ADC32736;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human novel contig-encoded polypeptide sequence, SEQ ID NO:2818.  
XX  
KW Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 9.  
XX  
OS Homo sapiens.  
XX  
PN WO2003029271-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 24-SEP-2002; 2002WO-US030474.  
XX  
PR 24-SEP-2001; 2001US-0324631P.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX  
DR WPI; 2003-371981/35.  
DR N-PSDB; ADC31969.  
XX  
PT New polynucleotide and polypeptide useful for diagnosing, preventing or  
PT treating conditions such as neurodegenerative diseases, anemias, platelet  
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
PT cancer.  
XX  
PS Example 2; SEQ ID NO 2818; 1185pp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
CC invention also relates to nucleic acid sequences over 99% identical with  
CC the novel human cDNAs. The invention additionally encompasses expression  
CC vectors and host cells comprising a nucleic acid of the invention; the  
CC recombinant production of a polypeptide of the invention; an antibody  
CC against a polypeptide of the invention; a method of detecting  
CC polynucleotides or polypeptides of the invention; and methods of  
CC identifying a compound which binds to a polypeptide of the invention. The  
CC invention further discloses methods of preventing, treating or  
CC ameliorating a medical condition; kits comprising polynucleotide probes  
CC and/or monoclonal antibodies for carrying out the methods of the  
CC invention; methods for the identification of compounds that modulate the  
CC expression or activity of the polynucleotide and/or polypeptide; and 767  
CC contig sequences corresponding to the cDNA sequences of the invention  
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
CC -ADC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a human contig-  
CC encoded polypeptide sequence used in an example of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 160 AA;

Query Match 17.8%; Score 685; DB 7; Length 160;  
Best Local Similarity 100.0%; Pred. No. 1.6e-59;  
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKLGELLHDKLFLFEAM 60  
Db 31 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKLGELLHDKLFLFEAM 90  
QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKDKLTLPDELIGIMDTCFCLITWL 120  
Db 91 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKDKLTLPDELIGIMDTCFCLITWL 150  
QY 121 EGHSLAQTVF 130  
Db 151 EGHSLAQTVF 160

RESULT 7  
AAB07449  
ID AAB07449 standard; protein; 733 AA.  
XX  
AC AAB07449;

XX 20-OCT-2000 (first entry)

XX Amino acid sequence of a murine tissue repair protein.

XX Tissue repair protein; orofacial clefting; wound healing; tissue repair.

OS Mus sp.

XX WO200040719-A2.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-GB0000003.

XX

PR 06-JAN-1999; 99GB-000000167.

XX (UYLE-) UNIV LEEDS.

PI Markham AF, Bonthron D;

XX WPI; 2000-465983/40.

DR N-PSDB; AAA58824.

XX New human and mouse nucleic acids encoding a tissue repair protein,  
PT useful for diagnosing and treating orofacial clefting, and for promoting  
PT wound healing and/or tissue repair.

PS Claim 22; Fig 8; 45pp; English.

XX The present sequence represents a murine tissue repair protein. Tissue  
CC repair gene polynucleotides are useful for determining expression of mRNA  
CC in selected target tissue, e.g. for diagnosing and treating orofacial  
CC clefting. They are also useful for determining the presence of DNA  
CC mutations in patients suffering from, or suspected to be suffering from  
CC orofacial clefting. The antibodies are also useful in the diagnosis of  
CC orofacial clefting. The polynucleotide is also useful for promoting wound  
CC healing and tissue repair

XX Sequence 733 AA;

Query Match 3.3%; Score 127; DB 3; Length 733;  
Best Local Similarity 20.9%; Pred. No. 0.0076;  
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

QY 217 EERDPEVEL---EHQOCLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAQKLM----- 265  
Db 67 EQLDGSLEYDNREEHAFVLRKDVLFSQLVETALLALGYSHSSA-AQAQGIKLGWRNP 125

QY 266 -----VQADLLSAIHNSLHHGIIQAQNDTTKGDHP-----IMMGFEPL--- 303  
Db 126 LPLSYVTDAPDATVADMLQDVYHVVTLKIQLOSCSKLEDLPAEQWNHATVRNALKELLE 185

QY 304 VNORLLPPTFPFYAKIKREEMVNYFA-----RLIDRIKTV-CEVVNLTNL--H 349  
Db 186 MNQSTLAKECPLSQSMISSIVNSTYYANVSATKQCFGRWYKKYKIKIVERENLSDY 245

QY 350 CILDFFCFEFSEQSPCVLSR-----SLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPV 404  
Db 246 CVLG-----QRPMLPNNQLASLGTNEQSPHSQIHSTPIRNQV-PALQPTMSPGL 297

QY 405 LSPK-----CYLYNN-----HQAKCID-SFVTH-----CVR--PFCSL 435  
Db 298 LSPQLSPQVLRQIAMAHLINQIAVSRLLAHPQAINQQFLNHPPIPRAVKPEPTNSS 357

QY 436 IQIHGHNRRQRDKLGHILEEFATLQDEFMTFFPNRAEKVDAALHTMLLKQEPQHQHLAC 495  
Db 358 VEVSPDIYQQVRDE---LKRASVSQAVFARVAFNRTQ---GLLSEILRKEEDPR----- 405

QY 496 LGTWVLYHNLRIQYLLSGFELE--LY-----SMHEYIYYIYWYLSE----- 535  
Db 406 TASQSLLVNLRAM-QNFINLPEVERDRYQDERERSMNPVSMVSSASSSSSSRTPOAK 464

QY 536 -----FLYAWLMSLTSRADGSQMAEERIMEEQKGRSSKTKKK 574  
Db 465 TSTPTTDLPIKVDGANVNITAAIYDEIQEQEKRAKVSQALFAKVAANKSQGLCELLRWK 524

QY 575 KKVRLPSREITMSQAYQNMCMAGMFKTMVAFDMGDKVRKPKFELD---SEQVRYEHRFAPF 631  
Db 525 ENPSPENRTL-----WENLC-----TIRRF-----LNLQHERDVIYEESRHHHSERMQ 569

QY 632 NSVMTTP-FVHYLQ-----FKEMSDLNKYSPPPQSPPELYVAASKHFQQAKMILENIP-- 682  
Db 570 HVVQLPPEPVQVLRHQSQOPTKESPPEEAPPPPPPTEDSCAKKPRSRTKISLEALGIL 629

QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703

Db 630 QSFIDVGLYPDQEAHTLSAQDLDPKHTTIK 661





QY 217 EERDPEVEL-----EHQQCLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAQKLM----- 265  
Db 95 EQLDGSLEYDNREEHAEFVLVRKDVLFSQLVETALLALGYSHSSA-AQAQGIILGRWNP 153  
QY 266 -----VQAADLLSAIHNSLHHGIIQAQNDTTKGDP-----IMMGFEPL--- 303  
Db 154 LPLSYVTDAPDATVADMLQDVYHVVTLKIQLSCKLEDLPAEQWNHATVRNALKELLKE 213  
QY 304 VNQRLLPPTFPFYAKIIKREEMVNYFA-----RLIDRIKTV-CEVNLTNL--H 349  
Db 214 MNQSTLAKECPLSQSMISSIVNSTYYANVSATKCOEFGRWYKKYKIKVERVERENLSY 273  
QY 350 CILDFCEFESEQSPCVLSR-----SLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPV 404  
Db 274 CVLG-----QRPMLPNMNQLASLGKTNEQSPHSQIHHSSTPIRNQV-PALQPI MSPGL 325  
QY 405 LSPK-----CYLYNN-----HQAQKCID-SFVTH-----CVR--PFCSL 435  
Db 326 LSPQLSPQLVRQIQIAMAHLINQIQIAVSRLLAHQHQPAINQQFLNHPPIPRAVKPEPTNSS 385  
QY 436 IQIHGHNRRQRDKLGHILEEFATLQDEFMTFYFNRAEKVDAAALHTMLLKQEPQOHLAC 495  
Db 386 VEVSPDIYQQVRDE-----LKRASVSQAVFARVAFNRTQ---GLLSEILRKEEDPR----- 433  
QY 496 LGTWVLYHNLRIMIQYLLSGFELE---LY-----SMHEYIYYWYLSE----- 535  
Db 434 TASQSLLVNLRAM-QNFLNLPEVERDRIYQDERERSMNPVSMVSSASSPSSSRTPQAK 492  
QY 536 -----FLYAWLMSTLSRADGSQMAEERIMEEQQKGRSSKTKKK 574  
Db 493 TSTPTDLPKVDGANINITAAIYDEIQQEMKRAKVSQALFAKVAANKSQGWLCCELLRWK 552  
QY 575 KKVRLPSREITMSQAYQNMCMGFKMTWAFDMGDKVKRKPKELD---SEQVRYEHRPAPF 631  
Db 553 ENPSPENRTL-----WENLC-----TIRRF-----LNLPHQERDVIYEEESRHHHSERMQ 597  
QY 632 NSVMTTP-PVHYLQ-----FKEMSDLNKYSPPPSPELYVAASKHFQQAAMILENIP-- 682  
Db 598 HVVQLPPEPVQVLRQQSQPAKESPPREEAPPPPTEDSCAKKPRSRTKISLEALGIL 657  
QY 683 -----NPDHEVNRIL--KVAKPNFVVMK 703  
Db 658 QSFHDVGLYPDQEAHTLSAQLDLPKHTTIK 689

RESULT 10  
ADD14191  
ID ADD14191 standard; protein; 761 AA.  
XX  
AC ADD14191;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE Human src biomarker polypeptide SEQ ID NO:380.  
XX  
KW predictor set; protein tyrosine kinase activity modulator;  
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytosstatic;  
KW gene therapy; drug sensitivity; genetic profile; cancer; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003062395-A2.  
XX  
PD 31-JUL-2003.  
XX  
PF 17-JAN-2003; 2003WO-US001981.  
XX  
PR 18-JAN-2002; 2002US-0350061P.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
FI Huang F, Fairchild CR, Lee FY, Shaw P;

XX  
DR  
DR  
XX  
PT  
PT  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

Query Match 3.3%; Score 127; DB 7; Length 761;  
Best Local Similarity 20.9%; Pred. No. 0.008;  
Matches 132; Conservative 91; Mismatches 227; Indels 182; Gaps 31;

QY 217 EERDPEVEL-----EHQQCLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAQKLM----- 265  
Db 95 EQLDGSLEYDNREEHAEFVLVRKDVLFSQLVETALLALGYSHSSA-AQAQGIILGRWNP 153  
QY 266 -----VQAADLLSAIHNSLHHGIIQAQNDTTKGDP-----IMMGFEPL--- 303  
Db 154 LPLSYVTDAPDATVADMLQDVYHVVTLKIQLSCKLEDLPAEQWNHATVRNALKELLKE 213  
QY 304 VNQRLLPPTFPFYAKIIKREEMVNYFA-----RLIDRIKTV-CEVNLTNL--H 349  
Db 214 MNQSTLAKECPLSQSMISSIVNSTYYANVSATKCOEFGRWYKKYKIKVERVERENLSY 273  
QY 350 CILDFCEFESEQSPCVLSR-----SLLQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPV 404  
Db 274 CVLG-----QRPMLPNMNQLASLGKTNEQSPHSQIHHSSTPIRNQV-PALQPI MSPGL 325  
QY 405 LSPK-----CYLYNN-----HQAQKCID-SFVTH-----CVR--PFCSL 435  
Db 326 LSPQLSPQLVRQIQIAMAHLINQIQIAVSRLLAHQHQPAINQQFLNHPPIPRAVKPEPTNSS 385  
QY 436 IQIHGHNRRQRDKLGHILEEFATLQDEFMTFYFNRAEKVDAAALHTMLLKQEPQOHLAC 495  
Db 386 VEVSPDIYQQVRDE-----LKRASVSQAVFARVAFNRTQ---GLLSEILRKEEDPR----- 433  
QY 496 LGTWVLYHNLRIMIQYLLSGFELE---LY-----SMHEYIYYWYLSE----- 535  
Db 434 TASQSLLVNLRAM-QNFLNLPEVERDRIYQDERERSMNPVSMVSSASSPSSSRTPQAK 492  
QY 536 -----FLYAWLMSTLSRADGSQMAEERIMEEQQKGRSSKTKKK 574

WPI; 2003-636735/60.  
N-PSDB; ADD14796.  
New polynucleotides and polypeptides for predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways.

Claim 10; SEQ ID NO 380; 139pp; English.

The present invention describes a predictor set comprising a plurality of polynucleotides or polypeptides whose expression pattern is predictive of the response of cells to treatment with a compound that modulates protein tyrosine kinase activity or members of the protein tyrosine kinase pathway. Also described: (1) predicting whether a compound is capable of modulating the activity of cells, comprising obtaining a sample of cells, determining whether the cells express a plurality of markers, and correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state, comprising subjecting the plurality of cell lines to one or more compounds, analysing the expression pattern of a microarray of polynucleotides or polypeptides, and selecting polynucleotides or polypeptides that predict the sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytostatic activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

Sequence 761 AA;







of the UPF1 and UPF3 genes (proteins Upflp and Upf3p, respectively) are essential components of this pathway. Mutations in these genes stabilise mRNA containing premature stop codons. The present sequence is the NMD2 gene protein from *Saccharomyces cerevisiae*: Nmd2p. The carboxy terminal of this protein binds to Upflp to inhibit its function. Inhibition of the nonsense-mediated mRNA decay pathway is a useful means of treating disorders caused by the presence of nonsense mutations, e.g. breast cancer, polycystic kidney diseases, Niemann-Pick disease, adenomatous polyposis coli, cystic fibrosis, Fanconi's anaemia, haemophilia, hypercholesterolaemia, neurofibromatosis, ornithine transcarbamylase deficiency, retinoblastoma, glycogen storage disease, McArdle disease, colorectal cancer, Tay-Sachs disease, Cowden disease, Wilson disease, polycystic kidney disease, Duchenne muscular dystrophy, adrenocortical carcinoma, Li-Fraumeni syndrome, Maple syrup urine disease, Turcot syndrome or beta-thalassaemia. Antisense mRNA molecules complementary to the NMD2 gene may be used for antisense therapy of the above mentioned disorders since the antisense molecules would inhibit translation of the NMD2 mRNA. Additionally, the NMD2 gene may be manipulated for ribozyme therapy of the above mentioned disorders: ribozymes designed to catalytically cleave nonsense-mediated mRNA decay pathway mRNAs e.g. NMD2 mRNA, can also be used to prevent translation and therefore expression of these mRNAs

SQ Sequence 1089 AA;

Query Match	3.3%;	Score 127;	DB 3;	Length 1089;
Best Local Similarity	17.6%;	Pred. No. 0.014;		
Matches 117;	Conservative 124;	Mismatches 202;	Indels 222;	Gaps 34;

QY	24	KSNTNWV-----DITQDFEEACRELKLGELL-----HDKXLFGLF	57
Db	35	KPNTGFIKKLKKGVKGSSESLKDLSEASLEKYLSEIIVTVECLLNVLNKNDVIAAV	94
QY	58	EAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIGDTIKIKDLTLPBLIGIMDTCECCL	116
Db	95	EIISGLHQRFNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR	136
QY	117	ITWLEGHSLAQTVFTCLYIHN-----PDFIE-----DPAMKAFALGILK-	155
Db	137	ITRVKGN--LRVFTELYGVFRTLDDIESKDAIPNFKQKTKGRKDPPLFSILREILNY	193
QY	156	-----ICDIAREKVNK-AAVFEED--FQSMTYGFKMANSVTDL-----RVTGM	196
Db	194	KFKLGFTTTIATAFIKKFAPLFRDDDNSWDDLIYDSKLKALQSFLKKNFIDATFARATEL	253
QY	197	LKDVEDDMQRRVKSTRSQGEERDPVELEHQOCLAVFSRVKFTRVLLTVLIAFTKKETS	256
Db	254	HKKV-NKLQREHQKCIQRTGKLRDEYVE-EYDKLLPIPIRFK-----TS	295
QY	257	AVAEAQKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPHIMMGFEPLVNQRLLPPTPPRY	316
Db	296	AITLGEFFKLEIPELEGASNDL-----KETA-----SPMITNQILPPN---	334
QY	317	AKIIKREEMVNYFARLIDRIKTVCV-----VNLTNLHCILDFFCFESEQSPCVL	366
Db	335	QRLWENEDTRKFYEILPDISKTVEESQSSKTEKDSNVNSKNINL---FFTDL-EMADC--	388
QY	367	SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA	417
Db	389	-----KDIIDDLNSRYWSSVLDNKATRNRILKFFMEFTQDWSKLPVYS-RFIATNSKYM	440
QY	418	KDCIDSFVTHCVRPFCSLIIQHGHNRARQDKLGHILEEFATLQDEFMTFFYNRAEKVDA	477
Db	441	PEIVSEFINYLDNGFRS--QLHS-----NKNVKNKI	469
QY	478	ALHTMLLKQEPORQHLACLGTVVLYHNLRLIMIQYLLSGFELELYSMHEYIYIYTWLSEFL	537
Db	470	IFFSEMIK-----FQLIPSEFIHFHKIRTLIMYMQVPPNNVEILTV-----LLEHSGKFL	517
QY	538	YAWLMSTLSRADGSQMAEERINBEEQKGRSSKTKTKKKKVRPLSREITMSQAYQNMCAGM	597
Db	518	-----LNKPEYKELMEKMV-----QLIKDKKNDRQLN--MNMKSALENIITLL	558

Db 8 RELREECRSLR-----SRRNQLERVSAME--DEMNMKREGKFRKRIKRNEQSLQEI 59

QY 94 TIKIKDLTLPELIGIMDTCECLITWLEGHSLAQVFTCLYIHNPDFIEDPAMKAFALGI 153

Db 60 WDYVKRPNL-RLIGVPES-----DAENGTKLENTL-----QDIIOEDFPNLARQAN 104

QY 154 LKICDIAR-----EKVNKAAVEEEDFQS----- 177

Db 105 VQIQEIQRTPQRYSSRRATPRHIIIVFTKVKEMQKMLRAA--REKDFKPTKIKRDKEGHY 162

QY 178 -MTYGFKMANSVTDLRV---TG---MLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQ 229

Db 163 IMVKGSIQQEELTILNIYAPNTGAPRFIKQVLSDLQRLDLSHTLIMGDFNTP----- 214

QY 230 CLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAAQKLMVQAADLLSAIHNSLHHG-----I 283

Db 215 -----LSILDRSTRQKVNKDTQELNSALHQAADLID-IYRTLHPKSTYEFF 259

QY 284 QAQNDT-TKGDHPIMMGFEPLVNQRLLPPTFFPRYAKIIKREEMVNYFARLIDRIKTVCEV 342

Db 260 SAPHHTYSKIDH--IVGSKALLSK-----CKRTQIITNYLSHSDHSAIKLELRI 304

QY 343 VNLTNLHCILDFCFESEQSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDP 402

Db 305 KNLTQ-----SRS---TTWKLNNMLNDYWVHNKMAEIKMFFET 342

QY 403 PVLSPKCYLYNNHQAQKCIDSFVTHCVRPFCSLIQIHGHNARQRDKLGH----- 453

Db 343 KENKDTTY-----QNLWDAFKAVCRGKF---IALNAHKRKQERSKIDTLTSQLKELEK 392

QY 454 -----LEEAFATLQ--DEFMTFFYNRAEKVDAAALHTMLLKQE 487

Db 393 QEQTNHSKASRRQEI TKIRAELEKEIETQKTLOKINESRSWFFERINKIDRPLARLIKKKR 452

QY 488 PQRQHLACLGTWVLYHNLIRIMIQYLLSG-----FELEYSMHEYIYYWYLSEFLYA 539

Db 453 EKNQ-----IDTIKNDKGDITTDPTTEIQT-TIREYY-----KHLYA 487

QY 540 WLMSTLSRADGSQMAEERIMEEQQGRSSKTKKKKKVVRPLSREITMSQAYQNMCAGMFK 599

Db 488 NKLENLEEMD-----KFLDITYTLPRLN-----QEEVESLNRPIDTSEI--EALINSLP 533

QY 600 TMVAFDMDGKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLOPKEMSDLNKYSPPPQ 659

Db 534 TKKSPGLDG-----FTAELYQ-RYKEELVPP-----LLKLFSIEKBRILEN 574

QY 660 SPELYVAASKHFQQAQMILENIPNPDHEVNRILKVAKPNFVVMKLL 705

Db 575 S-----FHEASITL--IPKPGRDTAK-----KENFRPISLM 603

RESULT 15  
ABG02338  
ID ABG02338 standard; protein; 1184 AA.

XX AC ABG02338;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #2329.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.  
XX PN WO200175067-A2.

XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS66525.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 20; SEQ ID NO 32697; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC involving aberrant protein expression or biological actions. The  
CC polypeptide and polynucleotide sequences have application in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1184 AA;

Query Match 3.2%; Score 121.5; DB 4; Length 1184;  
Best Local Similarity 20.6%; Pred. No. 0.057;  
Matches 123; Conservative 87; Mismatches 201; Indels 187; Gaps 31;

QY 171 BEEDFQSMTYGFKMANSVTDLRV---TG---MLKDVEDDMQRRVKSTRSRQGEERDPEV 223

Db 82 DKEGHYLMVKGSIQQEELTILNIYAPNTGAPRFIKQVLTDLQRLDLSHTLIMGDFNTP-- 139

QY 224 ELEHQQCLAVFSRVKFTRVLLTVLIAFTKKTSAVAEAAQKLMVQAADLLSAIHNSLHHG- 282

Db 140 -----LSTLDRSTRQKVNKDTQELNSALHQAADLID-IYRTLHPKS 178

QY 283 -----IQAQNDT-TKGDHPIMMGFEPLVNQRLLPPTFFPRYAKIIKREEMVNYFARLIDRI 336

Db 179 TEYTFFSAPHHTYSKIDH--ILGSKALLSK-----CQRTTEIITNYLSHDSAI 223

QY 337 KTVCEVVNLTNLHCILDFCFESEQSPCVLSRSLLOTTFLVDN-----KKVFGTHLMQD 390

Db 224 KLELRKLNLTQNH-----STTWKLNNLLNDYWVHNEMKAEIKMFFETNENKD 271

QY 391 M-----VKDALRSFVDPVLSPKCYLYNNH---QAKDCIDSFVTHCVRPFCSLIQIHGH-N 442

Db 272 TTYQNLDAFKA-----VCRGKFVALNAHKRKQERSKIDT-LTSQLKELEKQEQTSHKAS 325

QY 443 RARQRDKLGHILEEF---ATLQ--DEFMTFFYNRAEKVDAAALHTMLLKQEPQRQHLACLG 497

Db 326 RROQITKIGAELEKETQKTLOKINESRSWFFERINKIDRPLARLIKKKREKNQ----- 379

QY 498 TWVLYHNLIRIMIQYLLSG-----FELEYSMHEYIYYWYLSEFLYANLMSLTSRAD 549

Db 380 -----IDAIKNDKGDITTDPTTEIQT-TIREYY-----KHLYANKLENL---- 416

QY 550 GSQMAEERIMEEQQGRSSKTKK--KKKVRPLSREITMSQAYQNMCAGMFKTMVAFDMD 607



Db	417	-----EEMDKFLNTYTLPRLNQEEVESLNRPIITGSEIV--AIINSLRTKKSPGPD	464
Qy	608	GKVRKPKFELDSEQVRYEHRFAPFNSVMTPPVHYLQFKEMSDLNKYSPPQPSPELYVAA	667
Db	465	G-----FTAIFYQ-RYKEELVPF-----LLKQFQSIEKEGILPNS-----	498
Qy	668	SKHFQQAkmILENIPNPDHEVNRIILKVAKPNF-----VVMKLLAGGHKKESK	714
Db	499	---FYEASIIIL--IPKPGRDTTK-----KENFRPMSLMNIDAKILNKILANGIQQHIX	546

Search completed: August 10, 2004, 21:50:11  
Job time : 63 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:52:08 ; Search time 50 Seconds  
(without alignments)  
4598.592 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFFSAHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3849	100.0	733	13	US-10-001-857-145
2	3783	98.3	725	16	US-10-467-433-3
3	595.5	15.5	706	16	US-10-437-963-146818
4	322.5	8.4	256	12	US-10-424-599-190465
5	206.5	5.4	355	12	US-10-424-599-220965
6	139.5	3.6	965	12	US-10-403-571-40
7	136	3.5	2419	16	US-10-408-765A-2722
8	127	3.3	761	14	US-10-225-486-55
9	127	3.3	761	16	US-10-408-765A-696
10	127	3.3	969	12	US-10-335-977-7409
11	127	3.3	972	12	US-10-335-977-7410
12	121	3.1	745	12	US-10-425-114-65126
13	119	3.1	3353	9	US-09-888-615-64
14	116.5	3.0	1441	15	US-10-412-897-3
15	116.5	3.0	1641	12	US-09-964-956-40

16	116.5	3.0	1641	13	US-10-017-216-5	Sequence 5, Appli
17	116	3.0	736	12	US-10-170-385-285	Sequence 285, App
18	115	3.0	633	14	US-10-097-340-23	Sequence 23, Appl
19	115	3.0	1187	13	US-10-002-769-14	Sequence 14, Appl
20	115	3.0	1187	13	US-10-024-623-34	Sequence 34, Appl
21	115	3.0	1187	14	US-10-154-419-46	Sequence 46, Appl
22	115	3.0	1187	14	US-10-154-419-84	Sequence 84, Appl
23	115	3.0	1187	14	US-10-146-733-79	Sequence 79, Appl
24	114.5	3.0	892	16	US-10-437-963-200109	Sequence 200109,
25	114.5	3.0	1191	15	US-10-161-493-52	Sequence 52, Appl
26	114.5	3.0	1603	15	US-10-369-493-6744	Sequence 6744, Ap
27	114	3.0	1679	15	US-10-369-493-22080	Sequence 22080, A
28	113.5	2.9	917	16	US-10-437-963-200112	Sequence 200112,
29	113.5	2.9	1192	12	US-10-332-447-17	Sequence 17, Appl
30	113.5	2.9	1957	15	US-10-369-493-2070	Sequence 2070, Ap
31	112	2.9	679	16	US-10-437-963-175526	Sequence 175526,
32	112	2.9	1012	16	US-10-437-963-110215	Sequence 110215,
33	112	2.9	1786	9	US-09-742-096-3	Sequence 3, Appli
34	112	2.9	1787	12	US-10-415-253-2	Sequence 2, Appli
35	111.5	2.9	1332	13	US-10-041-856-3	Sequence 3, Appli
36	111.5	2.9	1332	13	US-10-041-856-5	Sequence 5, Appli
37	111	2.9	662	12	US-10-424-599-164683	Sequence 164683,
38	110	2.9	909	13	US-10-078-929-168	Sequence 168, App
39	110	2.9	909	14	US-10-078-770-190	Sequence 190, App
40	109.5	2.8	533	12	US-10-282-122A-61253	Sequence 61253, A
41	109.5	2.8	1321	16	US-10-437-963-191477	Sequence 191477,
42	109.5	2.8	1401	16	US-10-287-226-142	Sequence 142, App
43	109.5	2.8	2427	15	US-10-369-493-6734	Sequence 6734, Ap
44	109	2.8	794	15	US-10-320-797-3294	Sequence 3294, Ap
45	109	2.8	981	9	US-09-815-242-12211	Sequence 12211, A

ALIGNMENTS

RESULT 1

US-10-001-857-145  
; Sequence 145, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prote  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 145  
; LENGTH: 733  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-857-145

Query Match	100.0%;	Score 3849;	DB 13;	Length 733;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	733;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKLGELLHDKLFLFEAM	60	
Db	1	MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEACRELKLGELLHDKLFLFEAM	60	
Qy	61	SAIEMMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL	120	
Db	61	SAIEMMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL	120	
Qy	121	EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEEDFQSMTY	180	



Db 121 EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240  
Db 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240  
QY 241 RVLLTVLIAFTKKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPHIMMGF 300  
Db 241 RVLLTVLIAFTKKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPHIMMGF 300  
QY 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLDRIKTVCEVVNLTLNLHCILDFCFEFS 360  
Db 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLDRIKTVCEVVNLTLNLHCILDFCFEFS 360  
QY 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
Db 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFYFNRAEKVDAALH 480  
Db 421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFYFNRAEKVDAALH 480  
QY 481 TMLLKQEPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540  
Db 481 TMLLKQEPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540  
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600  
Db 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600  
QY 601 MVAFMDGKVRKPKFELDSEQVRYEHFAPNSVMTPPPVHYLQFKEMSDLNKYSPPQOS 660  
Db 601 MVAFMDGKVRKPKFELDSEQVRYEHFAPNSVMTPPPVHYLQFKEMSDLNKYSPPQOS 660  
QY 661 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Db 661 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
QY 721 FSAHKYFPVVKLV 733  
Db 721 FSAHKYFPVVKLV 733

RESULT 2

US-10-467-433-3  
; Sequence 3, Application US/10467433  
; Publication No. US2004008773A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION;  
; APPLICANT: LAL, Preeti G.; BAUGHN, Mariah R.;  
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;  
; APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;  
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;  
; APPLICANT: DING, Li; GIETZEN, Kimberly J.;  
; APPLICANT: ISON, Craig H.; LU, Dyung Aina M.;  
; APPLICANT: HAFALIA, April J.A.; GANDHI, Ameen R.;  
; APPLICANT: THANGAVELU, Kavitha; SANJANWALA, Madhusudan M.;  
; APPLICANT: TANG, Y. Tom; RAMKUMAR, Javalaxmi;  
; APPLICANT: GRIFFIN, Jennifer A.; SWARNAKAR, Anita;  
; APPLICANT: AZIMZAI, Yalda; SAPPERSTEIN, Stephanie K.;  
; APPLICANT: BURFORD, Neil; LEE, Ernestine A.;  
; APPLICANT: LU, Yan; TRAN, Uyen K.;  
; APPLICANT: MARQUIS, Joseph P.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
; FILE REFERENCE: PF-0899 USN  
; CURRENT APPLICATION NUMBER: US/10/467,433  
; CURRENT FILING DATE: 2003-08-06  
; PRIOR APPLICATION NUMBER: PCT/US02/03709  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/268,117  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/269,618  
; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: US 60/271,118  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: US 60/274,486  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 60/274,436  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: US 60/334,229  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: US 60/353,284  
; PRIOR FILING DATE: 2002-02-01  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PERL Program  
; SEQ ID NO 3  
; LENGTH: 725  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 7111920CD1  
US-10-467-433-3  
  
Query Match 98.3%; Score 3783; DB 16; Length 725;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 724; Conservative 0; Mismatches 1; Indels 8; Gaps 1;  
  
QY 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLFPEAM 60  
Db 1 MVMKASVDDDDSGWELSMPEKMEKSNNTNWVDITQDFEEACRELKLGELLHDKLFGLFPEAM 60  
QY 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120  
Db 61 SAIEEMDPKMDAGMIGNQVNRKVLNFEQAIDGTIKIKDITLPELIGIMDTFCCLITWL 120  
QY 121 EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180  
Db 121 EGHSLAQTVFTCLYIHNPDIEDPAMKAFALGILKICDIAREKVNKAAVFEEDFQSMY 180  
QY 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240  
Db 181 GFKMANSVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFT 240  
QY 241 RVLLTVLIAFTKKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPHIMMGF 300  
Db 241 RVLLTVLIAFTKKETSAAVAEAKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPHIMMGF 300  
QY 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLDRIKTVCEVVNLTLNLHCILDFCFEFS 360  
Db 301 EPLVNQRLPPTFPFYAKIIKREEMVNYFARLDRIKTVCEVVNLTLNLHCILDFCFEFS 360  
QY 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
Db 361 QSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLPSPKCYLYNNHQAKDC 420  
QY 421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFYFNRAEKVDAALH 480  
Db 421 IDSFVTHCVRPFCSLIIQHGNRRARQDKLGHILEEFATLQDEFMTFYFNRAEKVDAALH 472  
QY 481 TMLLKQEPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 540  
Db 473 TMLLKQEPQOHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFLYAW 532  
QY 541 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 600  
Db 533 LMSTLSRADGSQMAEERIMEEQKGRSSKTKKKKKVRPLSREITMSQAYQNMCMGFKT 592  
QY 601 MVAFMDGKVRKPKFELDSEQVRYEHFAPNSVMTPPPVHYLQFKEMSDLNKYSPPQOS 660  
Db 593 MVAFMDGKVRKPKFELDSEQVRYEHFAPNSVMTPPPVHYLQFKEMSDLNKYSPPQOS 652  
QY 661 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 720  
Db 653 PELYVAASKHFQQAAMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVPPEFD 712

QY 721 FSAHKYFPVVKLV 733  
Db 713 FSAHKYFPVVKLV 725

RESULT 3  
US-10-437-963-146818  
; Sequence 146818, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 146818  
; LENGTH: 706  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(706)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_47405C.1.pap  
US-10-437-963-146818

Query Match 15.5%; Score 595.5; DB 16; Length 706;  
Best Local Similarity 25.1%; Pred. No. 6.4e-48;  
Matches 191; Conservative 138; Mismatches 296; Indels 135; Gaps 25;

QY 25 SNTNWVDITQDFEEACRELKLGELLHDKLFGLEFAMSALFEMMDPKMDAGMIGNQVNRKVL 84  
Db 21 ASSAWADASPLDAAACRDLDQDELIGHGENFSLFGAMSALEIMDPKMDCGI----EKSGYY 76

QY 85 NFEQAIKDGTIKI-----KDLTLPELIGIMDTFCCLITWLEGHSLAQTFTCLYIHNPD 139  
Db 77 SIDEAIEDGIAPVPLSDRTLDIORTLDVMDHLFSCFATWHKGTHTLAQTFTCTIYLMRME 136

QY 140 FIEDPA-MKAFALGILKICDIAREKVNKAAVFEEEDFQSMTYGFKMANSVTDLRVTGMLK 198  
Db 137 RTSSHAILNSFCRILRATCYAVVSVVSTARTHEEDLFTMSFGLPLRDE-GDEKCLSI LN 195

QY 199 DVEDDMQRRVKSTRSR-----QG-EERDPEVELEHQOCLAVFSRVKFTFVLLTVLIA 249  
Db 196 SVEETIARQLRACKAQAALSRKKKTLEGLSLQDNPDLEEDYCRALLCRLRFRKHFFVHVTC 255

QY 250 FTKKETSAAVEAQK-----LMVQAADLLSAIHNSLHHGIAQND----TTKGDHP 295  
Db 256 LRKPHGRGLELAQKHVASCLTSLSLMNSRDLRSQSN-----TQQGDEICTTAGSVRP 311

QY 296 IMMGEPLVNQRLLPPTFPRYAKIIKREEMVNYEARLIDRIKTVCVVNLTLNHLCLDFF 355  
Db 312 V--GFDASLNSRLSPAPPRAVKLLSWSDAIRYFEKLLRDLDDICSSPLDPVLENVLHFV 369

QY 356 CEFSEQSPCVLSRSLLOTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNH 415  
Db 370 VQFQKSVPDLPVRAFLQ-TLLVQDGKLYGRDLSCDVISRALSL---PDIIGDKEFQMN-- 423

QY 416 QAKDCIDSFVTHCVRPFCSLIQIHGHNRARQDKLGHILEEFATLQDEFMTFFYFN--AE 473  
Db 424 -----EFVVQLGQLVINLLKILCTNTAWQRRKLGKSLQDMSTISIQ-LEFALKREFGE 475

QY 474 KVDAALH-TMLLKQEPQRQHLACLGTWVLYHNLRIMIQYLLSGFELELYSMHEYYIYWY 532

Db 476 TRNVLPHENCMRVSKQ-----LLVWTQEHYVWVAYRFLILGFELDLYSPSEYCMVY-- 527  
QY 533 LSEFLYAWLMSTLSRADGSQMAEERIMEEQQKGRSSKTKKKKKVRPLSREIT-----M 586  
Db 528 -----XRRKGK-----KKDHSKDSRRDTAFPSCLL 554

QY 587 SQAYQNMCMAGM-FKTMVAFDMDGKVR---KPKFELDSEQVRYEHRFAPFNSVMTPPPVHY 642  
Db 555 LQCYVLLSEGLSMDFQAAFSQXIHINSHWNMKLEFIALRGRFIOHFDLLQKARLPEHITY 614

QY 643 LQFKE-----MSDLNKYSPPPQSPELYVAASKHFQQAAMILENI-----PNPDHEVN 689  
Db 615 YSFRESASHASIADLTKYN-----FPKEIHKITPSLRGFSFASEPEKLAER 660

QY 690 RILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSAHKYFPV 729  
Db 661 QIEQVAEHNRIALNIISQVGAGDPSLRVSPFEFTHHPFAV 700

RESULT 4  
US-10-424-599-190465  
; Sequence 190465, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 190465  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(256)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_143009C.1.pap  
US-10-424-599-190465

Query Match 8.4%; Score 322.5; DB 12; Length 256;  
Best Local Similarity 31.8%; Pred. No. 3.2e-22;  
Matches 77; Conservative 49; Mismatches 93; Indels 23; Gaps 6;

QY 17 SMPEKM--EKSNTNMVVDITQDFEEACRELKLGELLHDKLFGLEFAMSALFEMMDPKMDAG 73  
Db 6 SLPPRASIPSNADNSVWADVSPLLQAACQDLOEGELIHGDNFNLFAMSALEIMDPKMDSG 65

QY 74 MIGNQVNRKVLNFEQAIKDGTIKI-----KDLTLPELIGIMDTFCCLITWLEGHSLAQT 128  
Db 66 -----IACTYYSLDEALENGVAVPVPIADKTTDVRCMIDIMDHLACEATWHKGHSLAQT 120

QY 129 VFTCLYIHNPDFIEDPA-MKAFALGILKICDIAREKVNKAAVFEEEDFQSMTYGFKMANS 187  
Db 121 VYSCLYLLRPERTSSHALLSYCKVIRATCKAILSVVSQARTHEEDLFTMAYGLPLSGD 180

QY 188 VTDLRVTGMLKDVEDDMQRRVKSTRSRQGEERDPE-----VELEHQOCLAVFSRVKF 239  
Db 181 -GDEKCLSLMNAVEETISRQLRACKVSSSKRRVPEDIEPLQNNPDLEEGYCKALXCRLRF 239

QY 240 TR 241  
Db 240 RK 241

RESULT 5





```

645  -----FKEMSDLN---KYSPPPQSPPELYVAAASKHFQQAAMILENI--- 681
      |:::  :  ||:::  :  ||:::  :  ||:::  :  ||:::  :
Db  779  DTVIAILKIPLSFORFYFFQKLQSTSIKALSPSPRTPEPIPQNN-QQLTLKVEGVIOH 837

QY  682  ---PNPDHEVNRI-LKVA-----KP--NFVVMKLLAGGHKKESKVPPEFDFS AHKY 726
      |:::  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :
Db  838  GSSPGLFRKIQAVCLKVSSTLQTKPGSDFKI-PLESKTNEIEQKVEPHNDYFSTQF 892

RESULT 7
US-10-408-765A-2722
; Sequence 2722, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2722
; LENGTH: 2419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2722

```

QY	467	FYF-----NRAEKVDAAL---HTMLLKQ--EPQRQH-----	492
Db	2048	FAHKASALNNWCKMEENLSEPVHCVSLNEIRQLQKDHEDFLASLARAQADFKLLELDQ	2107
QY	493	-LACLG-----TWVLYHNLIRIMIQYL-----LSGFEL-ELYSMHEY	526
		:	
Db	2108	QIKALGVSPSPYTWLTVEVLERTWKHLSDIIEEREQELQKEEARQVKNFEMCQFFEQNAS	2167
QY	527	YYIYWYLSEFLYAWLMSTLSRADGSQMAEERIMEEQOKGRSSKTKKKKKVRLSREIT-	585
		:	
Db	2168	TFLQWILETRAYF-----LDGSLKETGTLESQLEANK-----RKQKEIQAMKRQLTK	2215
QY	586	-----MSQAYQ---NMCAGMFKTMVAFMDGKVRKPKPFE	616
Db	2216	IVDLGNLEDALILDIKYSTIGLAQQWDQLYQLGLRMQHNLEQQIQAKDIKGVSEETLKE	2275
QY	617	LDSEQVRYEHRFAPFNSVMTPPPVHYLOFKE-MSDLNKYSPPQSPEL-----	663
		:	
Db	2276	FST---IYKH---FDENLTGRLTH-KEFRSCLRGLNYYLPMVEEDEHEPKFEKFLDAVD	2327
QY	664	-----YVAASKHFQQAQKMI---LENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKV	715
Db	2328	PGRKGYVSLDY--TAFLLDKESENIKSSDEIENAFQALAEKSYITK---EDMKQALT	2381
QY	716	PPEFDFA 723	
		:	
Db	2382	PEQVSFCA 2389	
RESULT 8			
US-10-225-486-55			
; Sequence 55, Application US/10225486			
; Publication No. US20030129207A1			
; GENERAL INFORMATION:			
; APPLICANT: Meagher, Madeleine Joy			
; APPLICANT: King, Gordon E.			
; APPLICANT: Secrist, Heather			
; APPLICANT: Jiang, Yuqiu			
; APPLICANT: Stolk, John A.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY			
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER			
; FILE REFERENCE: 210121.591			
; CURRENT APPLICATION NUMBER: US/10/225,486			
; CURRENT FILING DATE: 2002-10-29			
; NUMBER OF SEQ ID NOS: 67			
; SOFTWARE: Corixa Invention Disclosure Database			
; SEQ ID NO 55			
; LENGTH: 761			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-225-486-55			

	Query Match	3.3%; Score 127; DB 14; Length 761;	
	Best Local Similarity	20.9%; Pred. NO. 0.014;	
	Matches 132; Conservative	91; Mismatches 227; Indels 182; Gaps 31;	
QY	217 EERDPEVL----	EHQQC LAVFSRVKFT R V L L T V L I A F T K K T S A V A E A Q K L M-----	265
	:   :	:   :   :   :   :   :   :   :   :   :	
Dd	95 EQLDGSLEYDNREEHA	FVLRKDVLFSQLVETALLALGYSHSSA-AQAQGIIKLGRWNP	153
	:   :	:   :	
QY	266 -----	VQAADLLSAIHNSLHHGIGIAQNDDTTKGDPH-----IMMGEPL---	303
	:   :	:   :   :   :   :   :   :   :   :	
Dd	154 LPLSVYTDAPDATV	ADMQLQDVYHVVTLKIQLOSCSKLEDLP AEQWNHATVRNALKE LLKE	213
	:   :	:   :	
QY	304 VNQRLLPPTFPRYAKI	KREEMVNYFA-----RLIDRIKTVC EVNLTNL--H	349
	:  :    :	:  :    :	:
Dd	214 MNQSTLAKECPLSQS	MISSIVNSTYYANVSATKCQEFGRWYKKYKKIKIVERENLSDY	273
	:   :	:   :	
QY	350 CILDFCEFESEQSPCV	LRSR-----SLLQTTFLV DNKKVFEGTHLMQDMVKDALRSFVDPPV	404
	:   :	:   :   :   :   :   :   :   :   :	
Dd	274 CVLG-----	QRPMHLPNMNQLASLGKTNESPHSQIHHSTPIRNQV-PALQIPMSPG L	325
	:   :	:   :	



```
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...969
; SEQUENCE DESCRIPTION: SEQ ID NO: 7409:
US-10-335-977-7409

Query Match          3.3%; Score 127; DB 12; Length 969;
Best Local Similarity 17.3%; Pred. No. 0.021;
Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;

QY 35 DFEACRELKLGELLHDKLFLGFEAAMSDPKMDAGMIGNQV-----NRK 82
Db 9 DYQEQCR-----DQILGVFK---GIDLREPENDIQRIANPVFEIGAIDLLLENIE 56

QY 83 VLNFEQAIDGTIKIKDLTLPELIGIMDT-----CFCCCLITWLEGHSLAQVFTCLY--- 134
Db 57 NLRSKQKITQGSVGINQSLNCDI--LMETGTGKTFCFL-----ECVYALHK 100

QY 135 -IHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMYGFKMANSVTDLRV 193
Db 101 NYHLSKFIVLAPSNAILKGLVLSIEITRE-----FFKSEYSNTHL-- 140

QY 194 TGMKLDVEDDMQRRVKSTRSRQGEERDPEVELEHQQCLAVFSRVKFTRVLLTV----- 246
Db 141 -----ESYEDVERFILAS-----NHKCCVLVMTFSAFNKEKNTINQSCLEN 181

QY 247 --LIAFTKKTETSAVAEAQKLMV-----QAADLLSAIHNSLHHGIIQAQNDTTK 291
Db 182 TNLFNGAKSYMQUALASMRPIVIMDEPHRFLGDKTKKYLEQLNAL-ITLRFGATFKDDY-- 238

QY 292 GDHPIMMGFEPLVNQRLLPPTFPFYAKIIKREEMVNYFARLID----- 334
Db 239 --HNLIYALD---SKKAFDCALVKSIASVSGESDEYFLELKEANKKQNEATINYTTLEN 293

QY 335 -----RIKTVCEVNVLTNLHCILDFECFSEQSPCVLSRLIQTFLVDNKKVFGTHLMQ 389
Db 294 KIQSVKVTNHDNLGALTHISALEDYIVEKITKTEI---RFLNGENLLLDQKEPF-SHLE 349

QY 390 D-----MVKDALRSFVD-----PPVLSPKCYLYNNHQAOCIDSFVTHCVRP-----FC 433
Db 350 GEQVNLKEAIKSHFEREEGLFKKGIKALCMVFING-----VNSYLSENEKPAKLALFE 404

QY 434 SLIIQHGNRRARQDXLGHILEEFATLQDEFMTFYFNRAE---KVDAALHTMLLKQEPQ 489
Db 405 KLYQ-----QKLEEVKK---PLDENYRAYLERTKDAILKVHGGYFAXSKKESDE 451

QY 490 RQHLAGLTWVLYHNLRIMIQYLLSGPELELYSMHEYYYIYVYLSEFLYAW---LMSTL 545
Db 452 VQVIA-----LILKEKELLSFDSDL---RFIFSQWALQE---GWDNPNVMTIC 494

QY 546 SRADGSQMAEERIMEEQKGRSSKTKKKKKVR----- 578
Db 495 KlapSS---SNITKLQIIGRGLRLAVNDKGERITKEHADDFVNELVVIPOVEGDFVG 550

QY 579 PLSREITMSQAYQNMCA-----MFKTWVAFMDGKVRKPKFELDSEQV 622
Db 551 AIQGEISEHSLIKQVFSAAEELEKSGMVKKGYGVLFETLEGLGFGKTDDENFKLTLNQ- 609

QY 623 RYEHRAFPNVMTPPP-----VHYLQFKEMSDLNKYSPPQSPPELYVAAASKHFQQAQ 675
Db 610 -----NEFLKKEPELEKLDKDYLDPEKLKDFLK-----DRLIG---HFR--- 646

QY 676 MILENIPNDPHEVNRILKVAKNPFFVVMKLLAGGHKKESKVPPEFD 720
Db 647 -----VRNKNRERKTEKIKINKENFKKPETLWAGLNHQARIAAYID 686
```

RESULT 11

```
US-10-335-977-7410
; Sequence 7410, Application US/10335977
; Publication No. US2004005279A1
; GENERAL INFORMATION:
```

```
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7410:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...972
; SEQUENCE DESCRIPTION: SEQ ID NO: 7410:
US-10-335-977-7410
```

```
Query Match          3.3%; Score 127; DB 12; Length 972;
Best Local Similarity 17.3%; Pred. No. 0.021;
Matches 143; Conservative 125; Mismatches 271; Indels 286; Gaps 37;
```

```
QY 35 DFEACRELKLGELLHDKLFLGFEAAMSDPKMDAGMIGNQV-----NRK 82
Db 12 DYQEQCR-----DQILGVFK---GIDLREPENDIQRIANPVFEIGAIDLLLENIE 59

QY 83 VLNFEQAIDGTIKIKDLTLPELIGIMDT-----CFCCCLITWLEGHSLAQVFTCLY--- 134
Db 60 NLRSKQKITQGSVGINQSLNCDI--LMETGTGKTFCFL-----ECVYALHK 103

QY 135 -IHNPDFIEDPAMKAFALGILKICIDIAREKVNKAAVFEEDFQSMYGFKMANSVTDLRV 193
Db 104 NYHLSKFIVLAPSNAILKGLVLSIEITRE-----FFKSEYSNTHL-- 143

QY 194 TGMKLDVEDDMQRRVKSTRSRQGEERDPEVELEHQQCLAVFSRVKFTRVLLTV----- 246
Db 144 -----ESYEDVERFILAS-----NHKCCVLVMTFSAFNKEKNTINQSCLEN 184

QY 247 --LIAFTKKTETSAVAEAQKLMV-----QAADLLSAIHNSLHHGIIQAQNDTTK 291
Db 185 TNLFNGAKSYMQUALASMRPIVIMDEPHRFLGDKTKKYLEQLNAL-ITLRFGATFKDDY-- 241

QY 292 GDHPIMMGFEPLVNQRLLPPTFPFYAKIIKREEMVNYFARLID----- 334
Db 242 --HNLIYALD---SKKAFDCALVKSIASVSGESDEYFLELKEANKKQNEATINYTTLEN 296
```



QY	335	----	R I K T V C E V N L T N L H C I L D F F C F C E F S E Q S P C V L S R S L L Q T T F L V D N K K V F G T H L M Q	389
D b	297	K I Q S V K V K T H D N L G A L T H I S A L E D Y I V E K I T K T E I ---R F L N G F N L L D O Q E P F -S H L L E	352	
QY	390	D-----M V K D A L R S F V D-----P P V L S P K C Y L Y N N H Q A K D C I D S F V T H C V R P-----F C	433	
D b	353	G E Q E V M L K E A I K S H F E R E E G L F K G I K A L C M V F I N G-----V N S Y L S E N E K P A K L A L L F E	407	
QY	434	S L I Q I H G H N R A R Q D K L G H I L E E F A T L Q D E F M T F Y F N R A E----K V D A A L H T M L L K Q E P Q	489	
D b	408	K L Y Q-----O K L E E V L K---P L D E N Y R A Y L E R T K D A I L K V H G Y F A K S K K E S D E	454	
QY	490	R Q H L A C L G T W V L Y H N L R I M I Q Y L L S G P E L E L Y S M H E Y Y I Y W Y L S E F L Y A W----L M S T L	545	
D b	455	V Q V I A-----L I L K E K E L L S F D S D L-----R F I F S Q W A L Q E---G W D N P N V M T I C	497	
QY	546	S R A D G S Q M A E E R I M E E Q Q K R S S K K T K K K K V R-----	578	
D b	498	K L A P S S-----S N I T K L Q I G R G L R L A V N D K G E R I T K E H A D F D F V N E L V V I P Q V E G D F V G	553	
QY	579	P L S R E I T M S Q A Y Q N M C A G-----M F K T M V A F D M D G K V R K P K F E L D S E Q V	622	
D b	554	A I Q Q E I S E H S L I K Q V F S A E E L E K S G M V K K G Y Y G V L F E T L E G L G F G E K T D D E N F K L T L N Q-	612	
QY	623	R Y E H R F A P N S V M T P P P-----V H Y L Q F K E M S D L N K Y S P P P Q S P E L Y V A A S K H F Q Q A K	675	
D b	613	-----N E F L K K E P E L E K L D E K Y L D F E K L D F L K-----D R L I G-----H F R-----	649	
QY	676	M I L E N I P N P D H E V N R I L K V A K P N F V V M K L L A G G H K K E S K V P P E F D	720	
D b	650	-----V R N K N E R K T E K I K I N K E N F K F E T L W A G L N H O A R I A Y A I D	689	

RESULT 12

```

US-10-425-114-65126
; Sequence 65126, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)/B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65126
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17075F02_FLI.pep
US-10-425-114-65126

```

Query Match	3.1%;	Score 121;	DB 12;	Length 745;
Best Local Similarity	20.8%;	Pred. No. 0.052;		
Matches 129;	Conservative	91;	Mismatches 210;	Indels 190;
				Gaps 36;

```

QY 206 RRVKSTRSRQGEERDPEVELEHQQ-----CLAVFSRVKFTFRLTLTLIAFTKKE 254
Db 4 RRARSILRPTGEAEVPRSCSISSAAKSANQIRPVCWCAGAGVGRGHRITLGTATWATFAKPE 63

QY 255 TSAAVEAQKLM---VQAADLLSAIHNSLHHGIIQAQNDTTKGDHPIMMGF-EPLVNQRL 309
Db 64 -NALKRAEELIHVGQKQAA--LQALHDLITS--KRYRSWQKPLEKIMMYVELCVDLRK- 117

QY 310 PPTFPFYAKIIKREEMVNYFARLIDRIKTVCEVVNLTLNHCILDFFCFESESQSPCVLSRS 369
Db 118 ---GRFAK---DGLIQY-----RIVCQQVNVSSLEDVIKHFMQLSNEK---AEQA 158

```

QY	370	LLQTTFLVDNKKVFGTHLMQDM-	-----VKDALRSFV-----	DPBVLS-	-----	KC	409
Db	159	KSQVEALEDALDV-----	EDLEADKRPELDMLSFVSGEKGDRSDKEVVTPWFKFLWET	212			
QY	410	Y-----LYNNHQAKDCIDSFVTH-----	C-----VRPFCSLIQLIHGHNARQORDKLG	451			
Db	213	YRTVLEILRNNSKLEALYAMAHAHRAQFCQYKRTTEFRRLCEIIRHNLANLNKYRDQD	272				
QY	452	HILEEFATLQDEFMTFYFNRAEKVDAAALHTMLLKQEPQROHACLGTWVLYHNLRIMIQY	511				
Db	273	---RPDLTAPESLQLYLDRVEQLKVATE-LCLWQEAFR-----	SVEDIHGLMSMVK-	320			
QY	512	LLSGFELELYSMHEYYY-----IYWYLSEFL--	YAWLM-----STLSRADGSQMA	554			
Db	321	-----KMPKPSILVVVYAKLTETFPWVSDSHLYHAYAWLKLFLNLOKNYKNLSQKDLQLIA	375				
QY	555	EERIME-----EQQGRSSKTKKKKKVRPLSREITMSQAYQNMCAQMEKTMVAFDMD	607				
Db	376	SSVLLAALSVAPYDKKYGASHLETEKE-----	RNMRA-----NLVNFSLD	418			
QY	608	GK-----VRPKPFELDSEQVR-----	YEHRFAPENSVMTPPPVHYLQFK	646			
Db	419	SKRENREMPRASLLSELASKGVLSCASQEVDRDLYNLLLEHEFEPLDLASKVQPL-----	LS	474			
QY	647	BMSDL-NKYSPPPPQSPEL-----	YVAASKH-----FOQAKMILEN-----IPNPD-	685			
Db	475	KISKMGKGLSSASSVPEVKLSQYISALEKLTALRVLQQAASCIQSGMKIDMLSRMIPFFDF	534				
QY	686	HEVNRILKVA-KPNFVVMKL	704				
Db	535	FVVEKISVAVRRNFVAIKV	554				

RESULT 13

```

US-09-888-615-64
; Sequence 64, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 3353
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1891)
; OTHER INFORMATION: Any amino acid
US-09-888-615-64

```

Query Match	3.1%	Score 119;	DB 9;	Length 3353;
Best Local Similarity	19.3%	Pred. No. 0.89;		

QY	97	IKDLTLP	ELIGIM	DTCFCL	LITWLE	GHSLAQ	TVFTCL	YIHNPD	FIEDPA	KAFALGI	XI	156
		:::	:::	:	:	:	:	:	:	:	:	
Db	2720	MREELED	IKQFKK	TTSVLR	CIDGRS	CTTLIS	-----	AFRILLE	2761			
QY	157	CDIAREKV	--NKA	AVERBED	FQSMTY	GFKMANS	VTDL	VTGML	KDVEDD	MQRRVK	STRSR	214
					:	:	:	:	:	:	:	

Db	2762	SDEDRLVFNRLILMTESFNTL---HMVYEATACHVTGDLVELLSIFLSVLKSTRPY	2811
QY	215	QGEERDPEVELEHQOCLAVFSRVKFTRVLLTVLIAFTKKE-TSAVAEAQK--LMVQAADL	271
Db	2819	LQRKDVKQALIQWE-----RIEFAHKLLTLNLSYSPPELNRNACIDVLKELVLLSPHDF	2872
QY	272	LSAI-----HNSLHHGIIQAQNDTTKGDHPIMMG--FEPLVNQRL-----PP----	311
Db	2873	LHTLVFPLQHNHCTYHH-----SNIPMSLGPYFPCRENIKLGGKSNIRPPRPPEL	2922
QY	312	---TFPRYAKIIKREEMVNYFARLIDRIKTVEVNVNLNLHCILDFFCE-FSEQSPCVLS	367
Db	2923	NMCLLPFTWVETSKGDDV-YDRMLLDYFFSYHQFIHLL---CRVAINCEKTE-----	2971
QY	368	RSLLQTTFLVDNKKV-----FGTHLMQD---MVKDALRSFVDPVVLSP--KC-----	409
Db	2972	-TLVKLSVLVAYEGLPLHLALFPKLWTELCQTQSAMSKNCIKLLCEDPVFAEYIKCILMD	3030
QY	410	---LYNNHQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLG-----HILEEFATLQD	462
Db	3031	ERTFLNN-----IVYTFMTH-----FLLKVQSQVSEANCANLISTLITNLISQYQNLQS	3081
QY	463	EFMTFYFNRAE--KVDAALH-----TMLLKQBPQRQHLACLGTWLVYHNLRLIMIYLLS	514
Db	3082	DFS---NRVEISKASASLNGDLRALALLSVHTPKQ-----3114	
QY	515	GFELELYSMHEYIYIWYLSEFLYAWLMSTLSRADGSQMAERIMEEQKGRSSKTKKK	574
Db	3115	-----LNPALIPTLQELLSKCRCTLQQRNSLQEQEAKER---KTKDD	3153
QY	575	KKVRPLSREITMSQAYQNM--CAGMFKTMVAFDMDGKVRKPKFELDSEQVYEHRFAPFN	632
Db	3154	EGATPIKRRRRVSSDEEHTVDSCISDMKTETR-----EVLTPSTSDNE-----TRDS	3200
QY	633	SVMTPTPPVHYLOFKEMSDLNKYS---PPPQSPELYVAASKHFQQ	673
Db	3201	SIIDPGTEQDLSPENSSVKEYRMEVPSPSFSDMSNIRSQHAEE	3244

QY	671	FQQA	KMILENI-----PNPDHEVNRILKVAKPNFV-----VMKLLA--GG	708
Dd	691	LQKT	IELRSAREEAAHRAKATDHPHPSTPATARQQIAMSIVRSPHQPSAMSLAPSS	750
QY	709	HKKES	KVPPEF 719 :	
Dd	751	RRXES	TPEEF 761 :	
RESULT 15				
US-09-964-956-40				
; Sequence 40, Application US/09964956				
; Publication No. US20040043926A1				
; GENERAL INFORMATION:				
; APPLICANT: Gerlach, Valerie L				
; APPLICANT: MacDougall, John R				
; APPLICANT: Smithson, Glennnda				
; APPLICANT: Millet, Isabelle				
; APPLICANT: Stone, David				
; APPLICANT: Gunther, Erik				
; APPLICANT: Ellerman, Karen				
; APPLICANT: Grosse, William M				
; APPLICANT: Alsobrook II, John P				
; APPLICANT: Lepley, Denise M				
; APPLICANT: Burgess, Catherine E				
; APPLICANT: Padigaru, Muralidhara				
; APPLICANT: Kekuda, Ramesh				
; APPLICANT: Spytek, Kimberly A				
; APPLICANT: Leach, Martin D				
; APPLICANT: Shimkets, Richard A				
; TITLE OF INVENTION: No. US20040043926A1el Proteins and Nucleic Acids Encodin				
; FILE REFERENCE: 21402-124				
; CURRENT APPLICATION NUMBER: US/09/964,956				
; CURRENT FILING DATE: 2001-09-26				

---

RESULT 14				
US-10-412-897-3				
; Sequence 3, Application US/10412897				
; Publication No. US20030220224A1				
; GENERAL INFORMATION:				
; APPLICANT: Bristol-Myers Squibb Company				
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES ENCODING THE HUMAN CITRON KINASE				
; TITLE OF INVENTION: POLYPEPTIDE, BMSNKC_0020/0021				
; FILE REFERENCE: D0193 NP				
; CURRENT APPLICATION NUMBER: US/10/412,897				
; CURRENT FILING DATE: 2003-04-11				
; PRIOR APPLICATION NUMBER: U.S. 60/372,745				
; PRIOR FILING DATE: 2002-04-12				
; NUMBER OF SEQ ID NOS: 77				
; SOFTWARE: Patentin version 3.2				
; SEQ ID NO 3				
; LENGTH: 1441				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-10-412-897-3				
Query Match 3.0%; Score 116.5; DB 15; Length 1441;				
Best Local Similarity 19.0%; Pred. NO. 0.4;				
Matches 150; Conservative 132; Mismatches 316; Indels 193; Gaps 39;				
QY	20	EKMEKSNTNW	DITQDF---EEACRELKIGELLH---DKLFGLFEMSAIEMMDPKMDA	72
Dd	73	EKAVKASTEATE	LLQNIRQAKERAEREL---EKLNREDSSSEGIKKKLVEAELEEKHRE	129
QY	73	GMIGNQ----	VNRKVLNFEQAIK--DGTIKIKDLTLPELIGIMDTFCCLITWLEGH--	123
Dd	130	AQVSAQHLE	VHLVKQEQHYEEKIKVLDNQIK-KDLADKESLENM-----MQRHEE	178
QY	124	-----	SLAQTVTFTCLYIHNPDPFIEDBPAMKAFALGILKICDIAREKVNKAADVFEED	174





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 21:50:17 ; Search time 19 Seconds  
(without alignments)  
1991.676 Million cell updates/sec

Title: US-10-001-857-145  
Perfect score: 3849  
Sequence: 1 MVMKASVDDDDSGWELSMPE.....KVPPEFDFSAHKYFPVVKLV 733

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTus\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	3.3	1089	1 US-08-375-300-2	Sequence 2, Appli
2	127	3.3	1089	3 US-09-177-431-2	Sequence 2, Appli
3	127	3.3	1089	5 PCT-US95-16930-2	Sequence 2, Appli
4	113.5	2.9	1279	4 US-09-724-517-2	Sequence 2, Appli
5	113.5	2.9	1279	4 US-09-641-807A-2	Sequence 2, Appli
6	113.5	2.9	1279	4 US-09-723-096-2	Sequence 2, Appli
7	112	2.9	1786	3 US-08-973-462-8	Sequence 8, Appli
8	111.5	2.9	1332	2 US-08-971-244-2	Sequence 2, Appli
9	111.5	2.9	1332	3 US-09-286-891-2	Sequence 2, Appli
10	109.5	2.8	1066	3 US-09-541-782-8	Sequence 8, Appli
11	109.5	2.8	1066	4 US-09-723-820-8	Sequence 8, Appli
12	109.5	2.8	1066	4 US-10-270-085-8	Sequence 8, Appli
13	108	2.8	1388	2 US-08-685-576-1	Sequence 1, Appli
14	105.5	2.7	1507	4 US-09-914-259-37	Sequence 37, Appl
15	105	2.7	526	4 US-09-442-100-12	Sequence 12, Appl
16	105	2.7	526	4 US-08-939-106-12	Sequence 12, Appl
17	105	2.7	526	4 US-09-442-102-12	Sequence 12, Appl
18	104.5	2.7	585	4 US-09-107-532A-5494	Sequence 5494, Ap
19	104.5	2.7	1527	4 US-09-376-330-2	Sequence 2, Appli
20	102	2.7	732	2 US-08-533-669A-18	Sequence 18, Appl
21	102	2.7	732	4 US-09-307-143-4	Sequence 4, Appli
22	102	2.7	732	4 US-09-183-861-18	Sequence 18, Appl
23	102	2.7	732	4 US-09-022-765-18	Sequence 18, Appl
24	102	2.7	732	4 US-09-551-974A-18	Sequence 18, Appl
25	102	2.7	732	4 US-09-565-501A-18	Sequence 18, Appl
26	102	2.7	732	4 US-09-639-206A-18	Sequence 18, Appl
27	102	2.7	732	4 US-09-874-923-18	Sequence 18, Appl

28	101.5	2.6	1248	2 US-09-080-897-2	Sequence 2, Appli
29	101.5	2.6	1248	3 US-09-323-735-2	Sequence 2, Appli
30	101.5	2.6	1315	3 US-08-899-595-3	Sequence 3, Appli
31	101.5	2.6	2471	3 US-09-112-450-4	Sequence 4, Appli
32	101.5	2.6	2471	4 US-09-419-291A-4	Sequence 4, Appli
33	101.5	2.6	2662	4 US-09-595-684B-31	Sequence 31, Appl
34	101.5	2.6	3878	4 US-09-914-259-11	Sequence 11, Appl
35	100.5	2.6	2165	1 US-08-514-975B-2	Sequence 2, Appli
36	100.5	2.6	2165	5 PCT-US95-12507-2	Sequence 2, Appli
37	99	2.6	1581	4 US-09-866-108A-15754	Sequence 15754, A
38	98.5	2.6	540	4 US-09-588-995A-115	Sequence 115, App
39	98.5	2.6	4536	4 US-09-180-422B-27	Sequence 27, Appl
40	98.5	2.6	4563	4 US-09-108-006C-1	Sequence 1, Appli
41	98	2.5	1306	3 US-08-999-774A-13	Sequence 13, Appl
42	98	2.5	3038	1 US-08-450-332-2	Sequence 2, Appli
43	98	2.5	3038	2 US-08-637-640-2	Sequence 2, Appli
44	98	2.5	3038	3 US-09-004-406C-2	Sequence 2, Appli
45	97.5	2.5	2482	1 US-08-328-254-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-375-300-2  
; Sequence 2, Application US/08375300  
; Patent No. 5679566  
; GENERAL INFORMATION:  
; APPLICANT: Feng, He  
; APPLICANT: Jacobson, Allan S.  
; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN  
; TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mrNA DECAY FUNCTION  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,300  
; FILING DATE: 20-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. P.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 04020/046001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)542-5070  
; TELEFAX: (617)542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1089 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-375-300-2

Query Match 3.3%; Score 127; DB 1; Length 1089;  
Best Local Similarity 17.6%; Pred. No. 0.0033;  
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;  
Qy 24 KSNTNWV-----DITQDFEACRELKLGELL-----HDKLFGLF 57  
Db 35 KRNTGFIKKLKKGVKGSESSLKDLSEASLEKYLSEIIVTVTECLNLVINKNDDVIAAV 94

QY	58	BAMSAI-EMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLTLP	ELIGIMDT	CFCCCL	111
Db	95	ETISGLHQRFNGRFTSPLIG-----AFLQAFENPSVDIES-ERDEL-----	-----QR	136	
QY	117	ITWLEGHSLAQTVFTCLYIHN-----PDFIE-----	DPAMKAFALGILK-	155	
Db	137	ITRVKGN---LRVFTELYLVGVFRTLDDIESKDAIPNFKQKTGRKDP	LLFSILREILNY	193	
QY	156	-----ICDIAREKVNK-AAVFEED--FQSMTYGFKMANSVTDL-----	RVTGM	196	
Db	194	KFKLGFTTTIATAFIKKFAPLFRDDDNSDDLIYDSKLGALQSLFKNFIDATFARATEL	253		
QY	197	LKDVEDDMQRRVKSTRSQGEERDPEVELEHQQCLAVFSRVKFTRVLLTVLLI	FTKETS	256	
Db	254	HKKV-NKLQREHQKCQIRTKGLRDEYVE-EYDKLLPIFIRFK-----	TS	295	
QY	257	AVABAQKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDPIMMGFEPLVNQRL	LPPTFPY	316	
Db	296	AITLGEFFKLEIPELEGASNDLL-----KETA-----	SPMITNQILPPN----	334	
QY	317	AKIIKREEMVNYFARLIDRIKTVCEV-----VNLTNLHCILDFCFEFSEQSPCVL	366		
Db	335	QRLWENEDTRKFYEILPDISKTVESQSSTEKDSNVNSKNINL---FFIDL-EMADC--	388		
QY	367	SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----	PVLSPKCYLYNNHQA	417	
Db	389	-----KDIIDDLNRYWSSYLDNKATRNRIKFFMETQDWSKLPVYS-RFIATNSKYM	440		
QY	418	KDCIDSFVTHCVRPFCSLIQIHGHNRRQRDKLGHILEEFATLQDEFMTFYFNRAEKVDA	477		
Db	441	PEIVSEFINYLDNGFRS--QLHS-----	NKINVKNI	469	
QY	478	ALHTMLLKQEPQRQHLACLGTWVLYHNLIRIMIQYLLSGFELELYSMHEYYIYWLSEFL	537		
Db	470	IFFSEMIK-----FQLIPSEMIHFHKIRTLIMYMQVNNVEILTV-----	LLEHSGKFL	517	
QY	538	YAWLMSTLSRADGSGQMAEERIMEEQGRSSKTKKKKKVRPLSREITMSQAYQNMCAGM	597		
Db	518	-----LNKPEYKELMEXMV-----	QLIKDKKNDRQLN--MNMKSALENIITLL	558	
QY	598	FKTMV	602		
Db	559	YPPSV	563		

RESULT 2  
 US-09-177-431-2  
 ; Sequence 2, Application US/09177431  
 ; Patent No. 6071700  
 ; GENERAL INFORMATION:  
 ; APPLICANT: He, Feng  
 ; APPLICANT: Jacobson, Allan S.  
 ; TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE  
 ; TITLE OF INVENTION: ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: Windows 95  
 ; SOFTWARE: FastSEQ for Windows Version 2.0b  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/177,431  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/955,472  
 ; FILING DATE:

APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL  
TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE  
TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF  
TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECAY  
TITLE OF INVENTION: FUNCTION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street Suite 3100  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16930  
FILING DATE: 27-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/375,300  
FILING DATE: 20-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 04020/046WO1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)542-5070  
TELEFAX: (617)542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1089 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-16930-2

Query Match 3.3%; Score 127; DB 5; Length 1089;  
Best Local Similarity 17.6%; Pred. No. 0.0033;  
Matches 117; Conservative 124; Mismatches 202; Indels 222; Gaps 34;  
QY 24 KSNTNWV-----DITQDFEEACRELKLGELL-----HDKLFGLF 57  
Db 35 KRNTGFIKLLKGFVKGSESSLKDLSEASLEKYLSEIIVTVTECLNVLNKDDVIAAV 94  
QY 58 EAMSAI-EMMDPKMDAGMIGNQVNRKVLNPEQAIKDGTIKIKDLTLPDELIGIMDTFCCL 116  
Db 95 EIIISGLHQFPNGRFTSPLLG-----AFLQAFENPSVDIES-ERDEL-----QR 136  
QY 117 ITWLEGHSLAQTVFTCLYIHN-----PDFIE-----DPAMKAFALGILK- 155  
Db 137 ITRVKGN---LRVFTELYLVGVFRTLDDIESKDAIPNLFQKKTGRKDDPLLSILREILNY 193  
QY 156 -----ICDIAREKVNK-AAVFEED--FQSMYTYGFKMANSVTDL-----RVTGM 196  
Db 194 KFKLGFTTTTATAFIKKFPAPLFRDDNSWDLLIYDSKLKGALQSLFKNFIDATFARATEL 253  
QY 197 LKDVEDDMQRRVKSTRSRQGEERDPEVELEHQOCLAVFSRVKFRVLLTVLIAFTKKETS 256  
Db 254 HKKV-NKLQREHQKQIRTKGLRDEYVE-EYDKLLPIFIRFK-----TS 295  
QY 257 AVAAEQKLMVQAADLLSAIHNSLHHGIIQAQNDTTKGDHPIMMGFEPLVNOQLLPPTFPY 316  
Db 296 AITLGEFFKLEIPELEGASNDLL-----KETA-----SPMITNQILPPN---- 334  
QY 317 AKIIKREEMVNYFARLIDRIKTVCEV-----VNLTNLHCILDDFFCFEFSQSPCVL 366  
Db 335 QRLWENEDTRKFYEILDPDISKTVEESQSSKTEKDSNVNSKNINL---FFTDL-EMADC-- 388

QY 367 SRSLLQTTFLVD--NKKVFGTHLMQDMVKD-ALRSFVDP-----PVLSPKCYLYNNHQA 417  
Db 389 -----KDIIDDLNRYWSSYLDNKAATNRILKFFMETQDWSKLPVYS-RFIATNSKYM 440  
QY 418 KDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILBEFATLQDEFMTFYFNRAEKVDA 477  
Db 441 PEIVSEFINYLDNGFRS--QLHS-----NKNINVKNI 469  
QY 478 ALHTMLLKQEPQORQHLAGLTWVLYHNLRIMIQYLLSGFELELYSMHEYIYIYWLSEFL 537  
Db 470 IFFSEMIK-----FQLIPSEMFIFHKIRTLIMYQVNNVEILTV-----LLEHSGKFL 517  
QY 538 YAWLMSTLSRADGSQMAEERIMEEQOKGRSSKTKKKKKVVRPLSREITMSQAYQNMCAGM 597  
Db 518 -----LNKPEYKELMEKMV-----QLIKDKKNDRQLN--MNMKSALENIITLL 558  
QY 598 FKTMV 602  
Db 559 YPPSV 563  
RESULT 4  
US-09-724-517-2  
; Sequence 2, Application US/09724517  
; Patent No. 6379941  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6379941el motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1031  
; CURRENT APPLICATION NUMBER: US/09/724,517  
; CURRENT FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: US/09/641,807  
; PRIOR FILING DATE: 2000-08-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1279  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (409)...(436)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-724-517-2  
Query Match 2.9%; Score 113.5; DB 4; Length 1279;  
Best Local Similarity 20.3%; Pred. No. 0.092;  
Matches 155; Conservative 125; Mismatches 291; Indels 191; Gaps 39;  
QY 8 DDDDSGWELSMPEKMEKSNT-----NWVDITQDFEEACRELKLGELLHDKLFLGFAMSA 62  
Db 525 DDEES-----EGQEKSGTRCRRSRWI---QKPDVCSVLVELSDTQDE-----TQKSD 568  
QY 63 IEMMDPKMDAGMIGNQVN-RKVLNFEQAIKDGTIKIKDLT-----LPELIGI---- 108  
Db 569 LENEDLKIDCLQESQELNLQKLKNSERILTEAKQKMRELTINIKMKEDLIKELIKTGND 628  
QY 109 --MDTCFCCLITWLEGHS-----LAQTVFTCLYIHNPDFIEDPAMKAFALGILKICDIA 160  
Db 629 KSVSKQYSLKVTKLEHDAEQAKVELIETQKQLQELENKD-LSDVAMK-----VKLOKEF 681  
QY 161 REKVNKAAVFEEDFQSMYTYGFKMANSVTDLRVTMGLKOVDDMQRRVKSTRSRQGEERD 220  
Db 682 RKXMDAA-----KLRVQVLQKKQQDS---KKLASLSIQNEKRA 716  
QY 221 PEVE--LEHQOCLAVFSRVKFRVL-----LTVLIATFKKETSAAVAEAQ-----KLM 265  
Db 717 NELEQSVDDMK----YQKIQLQRKLRENEKRKQLDAVIKRDQQKIKELIKLTGQEEGLK 772  
QY 266 VQAADL-----LSAIHNSLHHGIIQAQNDTTKG-DHPIMMGFEPLVNOQ-----LLPPT 312



Db 773 PKAEDLDACNLKRRKGSFGSIDH-LQKLDQKKWLDEEV-----EKVLNQEQELELEADL 827

QY 313 FPRYAKIIKREEMVNYFARLIDRIKTVCEVNLTNLHC-----ILDFFCEFSQSPCVLS 367

Db 828 KXREAIVSKKEALLQEKSHLENKCLRSSQALNTDSLKISTRNLLE--QELSEKN----- 880

QY 368 RSLLOTTFVLVDNKKVFGTHLMQDMVKDAL---RSFVDDP-----VLSPK-----CYLYNN 414

Db 881 -VQLQTSTAEKTKISEQVEVLQKEKDQLQKRRHDVDEKLNKGRVLSPEEEHVLFLQLEEG 939

QY 415 HQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQD-EFMTF---YFN 470

Db 940 IEALEAAIEYRNESIQNRQKSLRASFNLSRGE---ANVLEKLACLSPVEIRTIILFRYFN 996

QY 471 RAEKVDAALHTMLLKQEPQOHL-----ACLGTVVLYHNLRIMIQYLLSGFELE 519

Db 997 KVVNLREAERKQQLYNEEMKMKVLERDNMVRLESALDHLKLQCDRRLTLQOKEHEQKMQ 1056

QY 520 LYSMHEYIYYWYLSEFLYAWLMSTL-SRADGSQMAEERIMEEQKGRSSKTKKKKKVVR 578

Db 1057 LLLHH-----FKEQDGEIMETFKTYEDKIQQLEKDLFYKKTSDRDKKKLKLVLGE 1108

QY 579 PLSREITMSQAYQNMCMAGMFKT---MVAFMDGKVRKPKF-----ELDSE----- 620

Db 1109 AIRRQLAPSE-YQEAGDGVLPKPEGGMLEELKWSRPESMKLSGREREMDSSASSLRTQ 1167

QY 621 ---QVRYEH--RFAPENSVMTPPPVHYL--QFKEMSDLNKYS 655

Db 1168 PNPQKLWEDIPELPPHSSSLAPPSGHMLGNENKTETDDNQFT 1209

RESULT 5

US-09-641-807A-2

; Sequence 2, Application US/09641807A

; Patent No. 6440731

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Freedman, Richard

; TITLE OF INVENTION: No. 6440731el motor proteins and methods for

; TITLE OF INVENTION: their use

; FILE REFERENCE: 1031

; CURRENT APPLICATION NUMBER: US/09/641,807A

; CURRENT FILING DATE: 2000-08-17

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1279

; TYPE: PRT

; ORGANISM: Human

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (409)...(446)

; OTHER INFORMATION: Xaa = any amino acid

US-09-641-807A-2

Query Match 2.9%; Score 113.5; DB 4; Length 1279;

Best local Similarity 20.3%; Pred. No. 0.092;

Matches 155; Conservative 125; Mismatches 291; Indels 191; Gaps 39;

QY 8 DDDDSGWELSMPEKMEKSNT-----NWVDITQDFEEACRELKLGELHDKLFLGLFEAMSA 62

Db 525 DDEES-----EGQEKSGTRCSRRSWI---QKPDVCSVLVELSDTQDE-----TKSD 568

QY 63 IEMDPKMDAGMIGNQVN-RKVLNFEQAIGDGIKIKDLT-----LPELIGI----- 108

Db 569 LENEDLKIDCLQESQELNLQKNSEIRILTEAKQKWRRELTINIKMKEDLIKELIKTGND 628

QY 109 --MDTCFCCLITWLEHGS-----LAQTVFTCLYIHNPDFIEDPAMKAFALGILKICDIA 160

Db 629 KSVSKQYSLKVTKLEHDAEQAKVELIETQKQLQELNKKD-LSDVAMK-----VKLQKEF 681

QY 161 REKVNKAAVFEEEDFQSMTYGFKMANSVTDLRVTGMKLDVEDDDMQRRVKSTRSQGEERD 220

Db 682 RKQMDAA-----KLRVQVLQKKQODS---KKLASLSIQNEKRA 716

QY 221 PEVE--LEHQQCLAVFSRVKPTRVL-----LTVLIAFTKETSAAVAEQ-----KLM 265

Db 717 NELEQSVDMHK---YQKIQLQKRLRENEKRRKQLDAVIKRDQKQIKELQKTGQEEGLK 772

QY 266 VQAADL-----LSAIHNSLHGIIQAQNDTTKG-DHPIMMGFEPLVNQR-----LLPPT 312

Db 773 PKAEDLDACNLKRRKGSFGSIDH-LQKLDQKKWLDEEV---EKVLNQEQELELEADL 827

QY 313 FPRYAKIIKREEMVNYFARLIDRIKTVCEVNLTNLHC-----ILDFFCEFSQSPCVLS 367

Db 828 KXREAIVSKKEALLQEKSHLENKCLRSSQALNTDSLKISTRNLLE--QELSEKN----- 880

QY 368 RSLLOTTFVLVDNKKVFGTHLMQDMVKDAL---RSFVDDP-----VLSPK-----CYLYNN 414

Db 881 -VQLQTSTAEKTKISEQVEVLQKEKDQLQKRRHDVDEKLNKGRVLSPEEEHVLFLQLEEG 939

QY 415 HQAKDCIDSFVTHCVRPFCSLIQIHGHNRRARQDKLGHILEEFATLQD-EFMTF---YFN 470

Db 940 IEALEAAIEYRNESIQNRQKSLRASFNLSRGE---ANVLEKLACLSPVEIRTIILFRYFN 996

QY 471 RAEKVDAALHTMLLKQEPQOHL-----ACLGTVVLYHNLRIMIQYLLSGFELE 519

Db 997 KVVNLREAERKQQLYNEEMKMKVLERDNMVRLESALDHLKLQCDRRLTLQOKEHEQKMQ 1056

QY 520 LYSMHEYIYYWYLSEFLYAWLMSTL-SRADGSQMAEERIMEEQKGRSSKTKKKKKVVR 578

Db 1057 LLLHH-----FKEQDGEIMETFKTYEDKIQQLEKDLFYKKTSDRDKKKLKLVLGE 1108

QY 579 PLSREITMSQAYQNMCMAGMFKT---MVAFMDGKVRKPKF-----ELDSE----- 620

Db 1109 AIRRQLAPSE-YQEAGDGVLPKPEGGMLEELKWSRPESMKLSGREREMDSSASSLRTQ 1167

QY 621 ---QVRYEH--RFAPENSVMTPPPVHYL--QFKEMSDLNKYS 655

Db 1168 PNPQKLWEDIPELPPHSSSLAPPSGHMLGNENKTETDDNQFT 1209

RESULT 6

US-09-723-096-2

; Sequence 2, Application US/09723096

; Patent No. 6448026

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Freedman, Richard

; TITLE OF INVENTION: No. 6448026el motor proteins and methods for

; TITLE OF INVENTION: their use

; FILE REFERENCE: 1031

; CURRENT APPLICATION NUMBER: US/09/723,096

; CURRENT FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: US/09/641,807

; PRIOR FILING DATE: 2000-08-17

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1279

; TYPE: PRT

; ORGANISM: Human

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (409)...(436)

; OTHER INFORMATION: Xaa = any amino acid

US-09-723-096-2

Query Match 2.9%; Score 113.5; DB 4; Length 1279;

Best local Similarity 20.3%; Pred. No. 0.092;

Matches 155; Conservative 125; Mismatches 291; Indels 191; Gaps 39;

QY 8 DDDDSGWELSMPEKMEKSNT-----NWVDITQDFEEACRELKLGELHDKLFLGLFEAMSA 62

Db 525 DDEES-----EGQEKSGTRCSRRSWI---QKPDVCSVLVELSDTQDE-----TKSD 568







Db 940 LKRYEKAIGHLSKCGPEYFPECLNLIKD-----KNLYNE--ALKLYSPSSQQ----- 984

QY 371 LQTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYLYNNHQAQKDCIDSFTV--HC 428

Db 985 -----YQDISIAYGEHLMQE-----HMYEPAGLMFARCGAH-----EKALSAFLT CGNW 1028

QY 429 VRPFCSLIQIH-----GHNRA-----RQRDKLGHILEEFATLQDEF----- 464

Db 1029 KQALCVAAQLNFTKDQLVGLGRTLAKGLVEQRKHIDAAMVLEECAQDYEEAVLLLEGA 1088

QY 465 -----MTFYFNRAEKVDAAALHTMLLKQEPQRQHLACLGTWVL---YHNLRIMIQYLLS- 514

Db 1089 WEEALRLVYKYNRLDIIETNVKPSIL--EAQKNYMAFLDSQTATFSRHKRLLVVRLEKE 1146

QY 515 -----GFELEYSMHEYIIYIYWYLSEFLYAWLMSTLSRADGSQM-----AE 555

Db 1147 QAQQAGLDDEVPHGQESDLFS-----ETSSVVGSEMSGKYSHSN 1186

QY 556 ERIMEEQQKGRSSKTKKKKK-----VRPLSREITMSQAYQONMCAGMFKTMV 602

Db 1187 SRISARSSKNR--RKAERKKHSLKEGSPLEDLALLEALSEVVQNTENLKDEVYHILKVL 1244

QY 603 AFDMD--GKVRKPKFELDSEQVRYEHRFAPFNSVMTPPPVHYLQFKEMSDLNKYSPP PQS 660

Db 1245 LFEFDEQGRELQKAFE-----DTLQLMERS-----L 1270

QY 661 PELYVAASKHFQQAQMILENIPNPDHEVNRILKVAKPNFVVMKLLAGGHKESKVP 716

Db 1271 PEIWTILT---YQQ-----NSATPVLGPNSTANSIMASYQQQKTSVP 1308

RESULT 10

US-09-541-782-8

; Sequence 8, Application US/09541782

; Patent No. 6284480

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Beraud, Christophe

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT APPLICATION NUMBER: US/09/541,782

; CURRENT FILING DATE: 2000-04-03

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 1066

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-541-782-8

Query Match 2.8%; Score 109.5; DB 3; Length 1066;

Best Local Similarity 19.7%; Pred. No. 0.17;

Matches 85; Conservative 80; Mismatches 168; Indels 99; Gaps 21;

QY 122 GH-SLAQTVFTCLYIH-----NPDFIEDPAMKAFALGILKI-CDIAREKVNKAA 168

Db 336 GHKDIEETLSTLEYAHRAKNIQNKPEVNQKLTTKTVLKEYTEEDKLRDLMAARDKNGI 395

QY 169 VFEEEDFQSMTYGFKMANSVTDLRVTGMKDVEDDMQRRVK--STRSRQGEERDPEVELE 226

Db 396 YLAETTYGEITLKLESQNRRELNEKML-LLKALKDELQNKKEKIFSEVSMSLVEKTQELKKT 454

QY 227 HQQCLAVFSRVKPTRVLLT-----VLIATTKKETS AVAEAKLMVQAADLLSAI 275

Db 455 EENLLNTKGTLLLTKKVLTTKRRYKEKKELVASHMKTEQVLT TQAEILA-AADLATDD 513

QY 276 HNSLHHGIAQNDDTKGDHPIMMGFEPLV-----NQRLLPPTFPY-----AKI KREE 324

Db 514 THQLHGTIERRREL---DEKIRRS CDQFKDRMQDNLEMIGGSLNLYQDQAAALKEQLSQE 570

QY 325 MVN-----YFARLIDRIKTV C--EVVNLTNLH-CILDDFFCFEFSQSPCVLSRSL 371

Db 455 EENLLNTKGTLLLTKKVLTTKRRYKEKKELVASHMKTEQVLT TQAEILA-AADLATDD 513

QY 276 HNSLHHGIAQNDDTKGDHPIMMGFEPLV-----NQRLLPPTFPY-----AKI KREE 324

Db 514 THQLHGTIERRREL---DEKIRRS CDQFKDRMQDNLEMIGGSLNLYQDQAAALKEQLSQE 570

QY 325 MVN-----YFARLIDRIKTV C--EVVNLTNLH-CILDDFFCFEFSQSPCVLSRSL 371

Db 571 MVNSSYVSQRLALNSSKSIEMLKEMCAQSLQDQTNLHNKLI GEVMKISDQ----- 620

QY 372 QTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYL-----YNNHQAQKDCIDS-- 423

Db 621 -----HSQAFVAKLMEQMQQQQL-----LMSKBIQTNLQVIEENNQRHKAMLD SMQ 666

QY 424 --FVTHCVRPFCSLIQIHGHNRRARQDKL GHI-----LEE FATLQDEFMTFYFNR--AEKV 475

Db 667 EK FAT-IIDSSLSQSV EEHAKQMHKKLEQLGAMSLPDABELQNLQEELAN---ERALAQQE 722

QY 476 DAALHTMLLKQE 487

Db 723 DALLESMMMQME 734

RESULT 11

US-09-723-820-8

; Sequence 8, Application US/09723820

; Patent No. 6468760

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Beraud, Christophe

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT APPLICATION NUMBER: US/09/723,820

; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: 09/541,782

; PRIOR FILING DATE: 2000-04-03

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 1066

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-723-820-8

Query Match 2.8%; Score 109.5; DB 4; Length 1066;

Best Local Similarity 19.7%; Pred. No. 0.17;

Matches 85; Conservative 80; Mismatches 168; Indels 99; Gaps 21;

QY 122 GH-SLAQTVFTCLYIH-----NPDFIEDPAMKAFALGILKI-CDIAREKVNKAA 168

Db 336 GHKDIEETLSTLEYAHRAKNIQNKPEVNQKLTTKTVLKEYTEEDKLRDLMAARDKNGI 395

QY 169 VFEEEDFQSMTYGFKMANSVTDLRVTGMKDVEDDMQRRVK--STRSRQGEERDPEVELE 226

Db 396 YLAETTYGEITLKLESQNRRELNEKML-LLKALKDELQNKKEKIFSEVSMSLVEKTQELKKT 454

QY 227 HQQCLAVFSRVKPTRVLLT-----VLIATTKKETS AVAEAKLMVQAADLLSAI 275

Db 455 EENLLNTKGTLLLTKKVLTTKRRYKEKKELVASHMKTEQVLT TQAEILA-AADLATDD 513

QY 276 HNSLHHGIAQNDDTKGDHPIMMGFEPLV-----NQRLLPPTFPY-----AKI KREE 324

Db 514 THQLHGTIERRREL---DEKIRRS CDQFKDRMQDNLEMIGGSLNLYQDQAAALKEQLSQE 570

QY 325 MVN-----YFARLIDRIKTV C--EVVNLTNLH-CILDDFFCFEFSQSPCVLSRSL 371

Db 571 MVNSSYVSQRLALNSSKSIEMLKEMCAQSLQDQTNLHNKLI GEVMKISDQ----- 620

QY 372 QTTFLVDNKKVFGTHLMQDMVKDALRSFVDPVLSPKCYL-----YNNHQAQKDCIDS-- 423

Db 621 -----HSQAFVAKLMEQMQQQQL-----LMSKBIQTNLQVIEENNQRHKAMLD SMQ 666

QY 424 --FVTHCVRPFCSLIQIHGHNRRARQDKL GHI-----LEE FATLQDEFMTFYFNR--AEKV 475

Db 667 EK FAT-IIDSSLSQSV EEHAKQMHKKLEQLGAMSLPDABELQNLQEELAN---ERALAQQE 722

QY 476 DAALHTMLLKQE 487

Db 723 DALLESMMMQME 734







Db 238 KHNYIHRD-----IKPDNLLL-----DRYGHKLKLSDFGLCKPLDCSTLEE 277

QY 463 BFMTFFYN-----RAEKVDAALHTMLLKQEPQHQHLACLGTWVLYHNLRIMIQY----- 511

Db 278 KDFSVDGNANGGSRSDSPAPKRT-----QQEQUEH-----WQKNRRMLAYSTVGTP 324

QY 512 -----LLSGFELELYSMHEYIYIYWYLSEFLYAWL-----MSTLSR-----A 548

Db 325 DYIAPEVLLKKGYGMEC-----DWWSLGAINYEMLVGYPPFYSDDPMSCTCRKIVNWK 376

QY 549 DGSQMAEERIMEEQQKG-----RSSKTKKKKKVRPLSREITMSQAYQNMCMAGM 597

Db 377 NHLKFPEEAKLSPEAKDIIISRLLCNVTERLGSNGADEIKVHSWFKGIDWDRIYQMEAA-- 434

QY 598 FKTWVAFDMDGKVRKPKFELDSEQVRYEHRFPFNSVMTPPVH-----YLOPKEMSD-- 650

Db 435 FIPEVNDELDTQ-NFEKFEESESHSQSGSRSGPWRKMLSSKDINFVGYYTKNFKVVDYQ 493

QY 651 -----LNKYSPPPPQSPEL 663

Db 494 VPGMVELKKTNTKPKKPTI 512

Search completed: August 10, 2004, 21:52:30

Job time : 23 secs